XITONG YU KONGZHI ZHONG DE JUZHEN LILUN

系统与控制中的

矩阵理论

张 显 仲光苹 高翔宇◇编著

黑龍江大學出版社

责任编辑：赵丽华
封面设计：张广东


ISBN 978-7-81129-403-3

定价：25.00元

XITONG YU KONGZHI ZHONG DE JUZHEN LILUN

系统与控制中的

矩阵理论

张　显　仲光苹　高翔宇◇编著

黑龍江大學出版社

图书在版编目(CIP)数据

系统与控制中的矩阵理论 / 张显，仲光苹，高翔宇编著．--哈尔滨：黑龙江大学出版社，2011.6

ISBN 978-7-81129-403-3

Ⅰ．①系… Ⅱ．①张… ②仲… ③高… Ⅲ．①矩阵-理论 Ⅳ．①O151.21

中国版本图书馆 CIP 数据核字(2011)第057259号

书　　名　系统与控制中的矩阵理论
著作责任者　张　显　仲光苹　高翔宇 编著
出 版 人　李小娟
责任编辑　赵丽华
出版发行　黑龙江大学出版社(哈尔滨市学府路74号　150080)
网　　址　http://www.hljupress.com
电子信箱　hljupress@163.com
电　　话　(0451)86608666
经　　销　新华书店
印　　刷　黑龙江省委党校印刷厂
开　　本　787×1092　1/16
印　　张　13.75
字　　数　230千
版　　次　2011年8月第1版　2011年8月第1次印刷
书　　号　ISBN 978-7-81129-403-3
定　　价　25.00元

前　言

本书的第一作者张显, 曾连续五年为黑龙江大学数学科学学院的研究生讲授"矩阵代数"课程, 本书是在其讲稿的基础上, 进行增删、改写而成的. 书中详细、准确地介绍了矩阵的列空间与核空间、矩阵(对)分解与标准形、向量范数、矩阵序列的极限与矩阵级数、函数矩阵的微积分、矩阵特征值和奇异值的不等式、矩阵广义逆、线性矩阵不等式、代数Riccati矩阵方程等方面的内容. 力求做到论述严谨、深入浅出, 能够体现矩阵的理论、思想和方法. 为了便于读者理解和消化内容, 每章均配有适量的习题, 较难的习题附有提示.

本书有两个特色: 一是为了内容的连贯、衔接, 方便读者阅读, 作者给出了许多与原文献不同的证明; 二是部分内容摘自近几年出版的控制方面的学术论文, 其中有的是作者的科研成果.

全书共分十二章, 第一至七章及附录A至D由仲光苹执笔, 第八至十一章及附录E和F由高翔宇执笔, 第十二章由张显执笔, 最后由张显统稿.

本书适合作为系统与控制等相关专业的研究生教材, 也可作为数学系本科生的选修课教材, 还可供相关专业的高等学校教师、广大科技工作者、工程技术人员参考.

本书的部分内容摘自国内外的同类著作、相关文献以及黑龙江大学曹重光教授的讲稿, 作者在此向这些作者表示衷心的感谢.

本书的出版得到了黑龙江省精品课程(近世代数)建设经费的资助, 在此深表谢意, 同时还要感谢黑龙江大学和东北石油大学的有关领导和同志的大力支持和帮助. 黑龙江大学数学科学学院的一些研究生参与了本书的打字和校对工作, 在此一并致谢.

由于我们水平有限, 本书可能有不当之处, 甚至错误, 热诚地欢迎各位同行专家和读者批评指正.

作者

2011年3月于哈尔滨

符号

如果没有特殊说明, 本书将使用下面的符号.

$\mathbb{R}$	实数集
$\mathbb{C}$	复数集
$\mathbb{F}$	$\mathbb{R}$或$\mathbb{C}$
$\mathbb{F}^n$	集合$\mathbb{F}$上所有n维列向量的集合
$\mathbb{F}^{m\times n}$	集合$\mathbb{F}$上所有$m\times n$矩阵的集合
$\mathbb{F}_k^{m\times n}$	集合$\mathbb{F}^{m\times n}$中所有秩为k矩阵的集合
$GL_n(\mathbb{F})$	集合$\mathbb{F}$上所有n阶可逆阵的集合
max	最大值
min	最小值
$\lambda_{\max}(\cdot)$	矩阵的最大特征值
$\lambda_{\min}(\cdot)$	矩阵的最小特征值
$\sigma_{\max}$	矩阵的最大奇异值
$\sigma_{\min}$	矩阵的最小奇异值
$\mathcal{R}(\cdot)$	矩阵的列空间
$\mathcal{N}(\cdot)$	矩阵的核空间
N_A	$\mathcal{N}(A)$的一组基作为列排成的矩阵
$\mathrm{span}\{\cdot\}$	生成子空间
$\mathrm{Im}\,(\cdot)$	线性变换的像空间
$\ker(\cdot)$	线性变换的核空间
E_{ij}	(i,j)位置元素为1而其余位置为0的矩阵
I_n或I	(n阶)单位阵
$0_{m\times n}$或0	($m\times n$阶)零矩阵
A_{ij}	矩阵A的(i,j)位置元素的代数余子式
$\mathrm{adj}(\cdot)$	伴随矩阵
$\det(\cdot)$	行列式
$\mathrm{rank}(\cdot)$	秩

$\lambda(\cdot)$	矩阵的所有特征根的集合
$\{\cdot\}_{o.n.}$	标准正交向量组
$A > B$或$B < A$	矩阵$A - B$是实对称(复Hermite)正定矩阵
$A \geqslant B$或$B \leqslant A$	矩阵$A - B$是实对称(复Hermite)半正定矩阵
$\rho(\cdot)$	谱半径
$\mathrm{tr}(\cdot)$	迹
$(\cdot)^{-1}$	逆
$(\cdot)^T$	转置
$(\cdot)^*$	共轭转置
$\overline{(\cdot)}$	共轭
$\|\cdot\|$	复数的模长
dim	维数
$\int$	积分
lim	极限
diag	对角矩阵
$(\cdot)^\perp$	正交补
$\\|\cdot\\|$	范数
$\\|\cdot\\|_1$	1-范数
$\\|\cdot\\|_2$	2-范数或谱范数
$\\|\cdot\\|_\infty$	∞-范数或H_∞范数
$\\|\cdot\\|_F$	Frobenius范数
$\mathrm{Cond}_{\\|\cdot\\|}$或Cond	条件数
sup	上确界
inf	下确界
$\mathrm{Ind}(\cdot)$	矩阵的指标
$(\cdot)_d$	矩阵的Draizin逆
$(\cdot)_g$	矩阵的群逆
$(\cdot)^{(1)}$	矩阵的$\{1\}$-逆
$(\cdot)^{\{1\}}$	矩阵的所有$\{1\}$-逆的集合
$(\cdot)^{(1,2)}$	矩阵的$\{1,2\}$-逆
$(\cdot)^{\{1,2\}}$	矩阵的所有$\{1,2\}$-逆的集合
$(\cdot)^{(1,3)}$	矩阵的$\{1,3\}$-逆
$(\cdot)^{\{1,3\}}$	矩阵的所有$\{1,3\}$-逆的集合
$(\cdot)^{(1,4)}$	矩阵的$\{1,4\}$-逆
$(\cdot)^{\{1,4\}}$	矩阵的所有$\{1,4\}$-逆的集合

$(\cdot)^+$	矩阵的Moore-Penrose逆
$\mu_{\|\cdot\|}(\cdot)$或$\mu(\cdot)$	矩阵测度
$\mathrm{Vec}(\cdot)$	拉直运算
M/A	矩阵M关于主子阵A的Schur补
$\widehat{M/A}$	矩阵M关于主子阵A的广义Schur补
$\frac{\mathrm{d}}{\mathrm{d}x}A$或$A'(x)$	变量A对变量x求导数
$\frac{\partial}{\partial x}A$	变量A对变量x求偏导数
$C_{k,X,Y}$	矩阵$\begin{bmatrix} Y & XY & \cdots & X^{k-1}Y \end{bmatrix}$
$P^{L,M}$	从子空间M到子空间L的投影算子
P^L	$P^{L,L}$
$P_{L,M}$	投影算子$P^{L,L}$在给定基下的矩阵
P_L	$P_{L,L}$
C_n^m	从n个元素中任取m个的组合数
$\mathbb{F}[x]_n$	$\mathbb{F}$上所有次数小于n的多项式和零多项式的集合
$\sum$	求和
$\prod$	乘积
$\otimes$	Kronecker积
$A^{[k]}$	$A \otimes A \otimes \cdots \otimes A$ (k个A)
$\circ$	Hadamard积
e_i	单位矩阵的第i列
$\mathrm{Re}(\cdot)$	实部
$\mathrm{Im}(\cdot)$	虚部
$\mathbb{C}^-$	左半开复平面
$\overline{\mathbb{C}^+}$	右半闭复平面
$:=$	定义
$\Leftrightarrow$	等价于
$\Rightarrow$	推出
$\forall$	任意的
$\subseteq$	包含于
$\supseteq$	包含
$\in$	属于
$\notin$	不属于
$\square$	结束符

目录

第1章　矩阵的列空间与核空间

矩阵的列空间和核空间是矩阵理论中的基本概念. 本章主要讨论矩阵的列空间和核空间的定义、性质以及它们的和是直和的充要条件.

1.1　矩阵的列空间与核空间的定义

若映射$\sigma:\mathbb{F}^n\to\mathbb{F}^m$满足

$$\sigma(ka+lb)=k\sigma(a)+l\sigma(b),\ \forall a,b\in\mathbb{F}^n,\ k,l\in\mathbb{F},$$

则称σ是从$\mathbb{F}^n$到$\mathbb{F}^m$的线性映射.

设$\varepsilon_1,\ \varepsilon_2,\ \cdots,\ \varepsilon_n$是$\mathbb{F}^n$的基, $\eta_1,\ \eta_2,\ \cdots,\ \eta_m$是$\mathbb{F}^m$的基, $\sigma:\mathbb{F}^n\to\mathbb{F}^m$是线性映射, 则$\sigma(\varepsilon_j)\in\mathbb{F}^m$, $j=1,2,\cdots,n$, 于是

$$\sigma(\varepsilon_j)=\sum_{i=1}^{m}a_{ij}\eta_i,\ j=1,2,\cdots,n,$$

其中$a_{ij}\in\mathbb{F}$. 任取$x=\sum\limits_{j=1}^{n}x_j\varepsilon_j\in\mathbb{F}^n$, 则

$$\begin{aligned}\sigma(x)=&\ \sum_{j=1}^{n}\left(x_j\sum_{i=1}^{m}a_{ij}\eta_i\right)\\=&\ \sum_{i=1}^{m}\left(\sum_{j=1}^{n}x_ja_{ij}\right)\eta_i\\=&\ \begin{bmatrix}\eta_1&\eta_2&\cdots&\eta_m\end{bmatrix}A_\sigma\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix},\ \forall x\in\mathbb{F}^n,\end{aligned}\tag{1.1.1}$$

其中

$$A_\sigma = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

由此容易得到下面的定理:

定理 1.1.1 对于给定的$\mathbb{F}^n$和$\mathbb{F}^m$的基, 映射$\sigma : \mathbb{F}^n \to \mathbb{F}^m$与$\mathbb{F}^{m\times n}$中的矩阵$A_\sigma$一一对应.

对于一个线性映射$\sigma : \mathbb{F}^n \to \mathbb{F}^m$, 称$\{x \in \mathbb{F}^n : \sigma(x) = 0\}$为映射$\sigma$的核空间, 记作$\ker\sigma$; 称$\{\sigma(x) : x \in \mathbb{F}^n\}$为映射$\sigma$的像空间, 记作$\mathrm{Im}\sigma$. 若取前面的$\eta_i$为单位矩阵$I_m$的第$i$列, ε_j为单位矩阵I_n的第j列, 则式(1.1.1)简化为

$$\sigma(x) = A_\sigma x, \ \forall x \in \mathbb{F}^n,$$

于是

$$\ker\sigma = \{x \in \mathbb{F}^n : A_\sigma x = 0\}, \quad \mathrm{Im}\,\sigma = \{A_\sigma x : x \in \mathbb{F}^n\}. \tag{1.1.2}$$

因而, 给出下面关于矩阵的核空间、列空间的定义是必要的.

定义 1.1.1 对于$A \in \mathbb{F}^{m\times n}$, 称$\{x \in \mathbb{F}^n : Ax = 0\}$ 为矩阵A的核空间, 记作$\mathcal{N}(A)$; 称$\{Ax : x \in \mathbb{F}^n\}$为矩阵$A$的列空间, 记作$\mathcal{R}(A)$.

从上面的定义易见, $\mathcal{N}(A)$和$\mathcal{R}(A)$分别是$\mathbb{F}^n$和$\mathbb{F}^m$的子空间, 并且

$$\dim\mathcal{N}(A) = n - \mathrm{rank}A, \ \dim\mathcal{R}(A) = \mathrm{rank}A.$$

这和式(1.1.2)一起可推出

$$\dim\ker\sigma = n - \mathrm{rank}A_\sigma,$$

$$\dim\mathrm{Im}\,\sigma = \mathrm{rank}A_\sigma.$$

1.2 列空间与核空间的性质和应用

本节介绍矩阵的列空间和核空间的一些常用的性质, 并给出它们在证明矩阵秩的不等式中的应用.

性质 1.2.1 设$A \in \mathbb{F}^{m\times n}$, $x_1, \cdots, x_r$是$\mathcal{N}(A)$的基, 将其扩充为$\mathbb{F}^n$的基$x_1, \cdots, x_r, x_{r+1}, \cdots, x_n$, 则$Ax_{r+1}, \cdots, Ax_n$是$\mathcal{R}(A)$的基.

证明 设$l_{r+1}Ax_{r+1}+\cdots+l_nAx_n=0$, 则$A(l_{r+1}x_{r+1}+\cdots+l_nx_n)=0$, 于是

$$l_{r+1}x_{r+1}+\cdots+l_nx_n\in\mathcal{N}(A).$$

因为$x_1,\cdots,x_r$是$\mathcal{N}(A)$的基, 所以存在$l_1,\cdots,l_r\in\mathbb{F}$, 使得

$$l_{r+1}x_{r+1}+\cdots+l_nx_n=l_1x_1+\ldots+l_rx_r.$$

由$x_1,\cdots,x_r,x_{r+1},\cdots,x_n$是$\mathbb{F}^n$的基推出

$$l_{r+1}=\cdots=l_n=0,$$

因而$Ax_{r+1},\cdots,Ax_n$线性无关.

任取$y\in\mathcal{R}(A)$, 则存在$z\in\mathbb{F}^n$, 使得

$$y=Az.$$

既然$x_1,\cdots,x_r,x_{r+1},\cdots,x_n$是$\mathbb{F}^n$的基, 可设$z=k_1x_1+\cdots+k_rx_r+k_{r+1}x_{r+1}+\cdots+k_nx_n$, 其中$k_i\in\mathbb{F}$, $i=1,2,\cdots,n$, 于是

$$\begin{aligned}y=Az &= k_1Ax_1+\cdots+k_rAx_r+k_{r+1}Ax_{r+1}+\cdots+k_nAx_n\\ &= k_{r+1}Ax_{r+1}+\cdots+k_nAx_n,\end{aligned}$$

即y可由$Ax_{r+1},\cdots,Ax_n$线性表出.

总之, $Ax_{r+1},\cdots,Ax_n$是$\mathcal{R}(A)$的基. □

性质 1.2.2 若$A\in\mathbb{F}^{n\times n}$, 则$\mathcal{N}(A)$和$\mathcal{R}(A)$均是$A$的不变子空间.

性质 1.2.3 设$A,B\in\mathbb{F}^{m\times n}$, 则

$$\mathcal{R}([A\quad B])=\mathcal{R}(A)+\mathcal{R}(B),\ \mathcal{R}(A+B)\subseteq\mathcal{R}([A\quad B]).$$

性质 1.2.4 设$A,B\in\mathbb{F}^{m\times n}$, 满足$\mathcal{N}(A)+\mathcal{N}(B)=\mathbb{F}^n$, 则

$$\mathcal{R}(\begin{bmatrix}A & B\end{bmatrix})\subseteq\mathcal{R}(A+B).$$

证明 任取$z\in\mathcal{R}(\begin{bmatrix}A & B\end{bmatrix})$, 则存在$x,y\in\mathbb{F}^n$, 使得$z=Ax+By$. 由$\mathcal{N}(A)+\mathcal{N}(B)=\mathbb{F}^n$可设

$$x=x_1+x_2,\ y=y_1+y_2,\ x_1,y_1\in\mathcal{N}(A),\ x_2,y_2\in\mathcal{N}(B),$$

于是$z=A(x_1+x_2)+B(y_1+y_2)=Ax_2+By_1=(A+B)x_2+(A+B)y_1\in\mathcal{R}(A+B)$. 故$\mathcal{R}(\begin{bmatrix}A & B\end{bmatrix})\subseteq\mathcal{R}(A+B)$. □

命题 1.2.1 设$A, B \in \mathbb{F}^{m\times n}$, 则

$$\operatorname{rank}(A+B) \leqslant \operatorname{rank}A + \operatorname{rank}B - \dim(\mathcal{R}(A)\cap\mathcal{R}(B)).$$

证明 由性质1.2.3得

$$\begin{aligned}
&\operatorname{rank}(A+B)\\
=\ &\dim\mathcal{R}(A+B)\\
\leqslant\ &\dim\mathcal{R}\left(\begin{bmatrix} A & B \end{bmatrix}\right)\\
=\ &\dim(\mathcal{R}(A)+\mathcal{R}(B))\\
=\ &\dim\mathcal{R}(A)+\dim\mathcal{R}(B)-\dim(\mathcal{R}(A)\cap\mathcal{R}(B))\\
=\ &\operatorname{rank}A+\operatorname{rank}B-\dim(\mathcal{R}(A)\cap\mathcal{R}(B)).
\end{aligned} \tag{1.2.1}$$

□

命题 1.2.2 设$A, B \in \mathbb{F}^{m\times n}$满足$\operatorname{rank}(A+B) = \operatorname{rank}A + \operatorname{rank}B$, 则

(i) $\mathcal{R}(A)\cap\mathcal{R}(B) = \{0\}$;

(ii) $\mathcal{N}(A)\cap\mathcal{N}(B) = \mathcal{N}(A+B)$;

(iii) $\mathcal{N}(A)+\mathcal{N}(B) = \mathbb{F}^n$.

证明 (i) 由命题1.2.1及$\operatorname{rank}(A+B) = \operatorname{rank}A+\operatorname{rank}B$推出$\dim(\mathcal{R}(A)\cap\mathcal{R}(B)) = 0$, 于是$\mathcal{R}(A)\cap\mathcal{R}(B) = \{0\}$.

(ii) 任取$x \in \mathcal{N}(A)\cap\mathcal{N}(B)$, 则$x \in \mathcal{N}(A)$且$x \in \mathcal{N}(B)$, 于是$Ax = 0$且$Bx = 0$, 从而$(A+B)x = 0$, 即$x \in \mathcal{N}(A+B)$. 故$\mathcal{N}(A)\cap\mathcal{N}(B) \subseteq \mathcal{N}(A+B)$.

任取$y \in \mathcal{N}(A+B)$, 则$(A+B)y = 0$, 于是$Ay = -By$. 由(i)推出$Ay = By = 0$, 进而$y \in \mathcal{N}(A)$且$y \in \mathcal{N}(B)$, 即$y \in \mathcal{N}(A)\cap\mathcal{N}(B)$. 故$\mathcal{N}(A)\cap\mathcal{N}(B) \supseteq \mathcal{N}(A+B)$.

(iii) 由(ii)推出

$$\begin{aligned}
&\mathcal{N}(A)+\mathcal{N}(B) = \mathbb{F}^n\\
\Leftrightarrow\ &\dim(\mathcal{N}(A)+\mathcal{N}(B)) = n\\
\Leftrightarrow\ &\dim\mathcal{N}(A)+\dim\mathcal{N}(B)-\dim(\mathcal{N}(A)\cap\mathcal{N}(B)) = n\\
\Leftrightarrow\ &n-\dim\mathcal{R}(A)-\dim\mathcal{R}(B)-\dim\mathcal{N}(A+B) = 0\\
\Leftrightarrow\ &\dim\mathcal{R}(A)+\dim\mathcal{R}(B) = \dim\mathcal{R}(A+B)\\
\Leftrightarrow\ &\operatorname{rank}(A+B) = \operatorname{rank}A+\operatorname{rank}B.
\end{aligned}$$

□

定理 1.2.1 设$A, B \in \mathbb{F}^{m\times n}$, 则

$$\operatorname{rank}(A+B) \leqslant \operatorname{rank}A + \operatorname{rank}B. \tag{1.2.2}$$

进而, 不等式(1.2.2)的等号成立的充要条件是

$$\mathcal{R}(A) \cap \mathcal{R}(B) = \{0\},\ \mathcal{N}(A) + \mathcal{N}(B) = \mathbb{F}^n. \tag{1.2.3}$$

证明 由命题1.2.1知式(1.2.2)成立. 注意到命题1.2.2, 只需证: 若式(1.2.3)成立, 则$\operatorname{rank}(A+B) = \operatorname{rank}A + \operatorname{rank}B$. 使用式(1.2.1)和性质1.2.4得证. □

设$A \in \mathbb{F}^{m\times n}, B \in \mathbb{F}^{n\times p}$, 则$AB \in \mathbb{F}^{m\times p}$, 若将$B$和$A$分别看成映射$B: \mathbb{F}^p \to \mathbb{F}^n$, $A: \mathbb{F}^n \to \mathbb{F}^m$, 则$AB$可看成映射$AB: \mathbb{F}^p \to \mathbb{F}^m$, 且$A$在$\mathcal{R}(B)$上的限制映射为$A|_{\mathcal{R}(B)}: \mathcal{R}(B) \to \mathbb{F}^m$, 从而

$$\begin{aligned}
\dim \operatorname{Im}(AB) &= \dim \operatorname{Im}(A|_{\mathcal{R}(B)}) \\
&= \dim \operatorname{Im}(B) - \dim(\operatorname{Im}(B) \cap \ker(A)) \\
&\geqslant \dim \operatorname{Im}(B) - \dim \ker(A) \\
&= \dim \operatorname{Im}(B) - (n - \dim \operatorname{Im}(A)) \\
&= \dim \operatorname{Im}(B) + \dim \operatorname{Im}(A) - n.
\end{aligned}$$

由此易见下面的定理成立.

定理 1.2.2 设$A \in \mathbb{F}^{m\times n}$, $B \in \mathbb{F}^{n\times p}$, 则

$$\operatorname{rank}(AB) \geqslant \operatorname{rank}A + \operatorname{rank}B - n. \tag{1.2.4}$$

进而, 不等式(1.2.4)的等号成立的充要条件是

$$\dim(\mathcal{R}(B) \cap \mathcal{N}(A)) = \dim \mathcal{N}(A)(\text{即}\mathcal{N}(A) \subseteq \mathcal{R}(B)).$$

注记 1.2.1 参考文献[1]从矩阵分解的角度给出了式(1.2.2)和(1.2.4)中等号成立的充要条件, 具体内容见3.4.

1.3 列空间与核空间的和是直和的条件

对于$A \in \mathbb{F}^{n\times n}$, 前面已经提到$\mathcal{N}(A)$和$\mathcal{R}(A)$均是$A$的不变子空间. 设$x_1$, x_2, $\cdots$, x_r是$\mathcal{R}(A)$的基, x_{r+1}, $\cdots$, x_n是$\mathcal{N}(A)$的基. 若$\mathcal{N}(A) \oplus \mathcal{R}(A) = \mathbb{F}^n$, 则$x_1$, x_2, $\cdots$, x_r, x_{r+1}, $\cdots$, x_n是$\mathbb{F}^n$的基, 于是

$$P^{-1}AP = \operatorname{diag}(A_1, 0),$$

其中$A_1 \in \mathbb{F}^{r\times r}$可逆, $P = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$. 总之, 当$\mathcal{N}(A) \oplus \mathcal{R}(A) = \mathbb{F}^n$时, 矩阵$A$将相似于一个形式较简单的矩阵, 这给处理问题带来很多方便. 因而我们关心$\mathcal{N}(A) \oplus \mathcal{R}(A) = \mathbb{F}^n$成立的条件.

命题 1.3.1 设$A \in \mathbb{F}^{n\times n}$, 则$\mathcal{N}(A) \oplus \mathcal{R}(A) = \mathbb{F}^n$的充要条件是

$$\mathcal{N}(A) + \mathcal{R}(A) = \mathbb{F}^n. \tag{1.3.1}$$

证明 必要性是显然的; 充分性由$\dim \mathcal{N}(A) + \dim \mathcal{R}(A) = n$得到. □

上面的命题表明要使矩阵A相似于一个较简单的形式, 只需式(1.3.1)成立. 然而, 下面的例子表明式(1.3.1)并不是对$\mathbb{F}^{n\times n}$ 中的所有矩阵都成立.

例 1.3.1 记$\mathbb{F}[x]_n$为$\mathbb{F}$上的所有次数小于n的多项式和零多项式的集合. 定义从$\mathbb{F}[x]_n$到$\mathbb{F}[x]_n$的线性映射$\sigma : f(x) \mapsto f'(x)$, $\forall f(x) \in \mathbb{F}[x]_n$. 取$\mathbb{F}[x]_n$的基$1, x, \frac{x^2}{2!}, \cdots, \frac{x^{n-1}}{(n-1)!}$, 则$\sigma$在这组基下的矩阵为

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{F}^{n\times n}.$$

易见

$$\mathcal{N}(A) = \left\{ \begin{bmatrix} x & 0 & \cdots & 0 \end{bmatrix}^T : x \in \mathbb{F} \right\},$$

$$\mathcal{R}(A) = \left\{ \begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} & 0 \end{bmatrix}^T : x_i \in \mathbb{F}, i = 1, 2, \cdots, n-1 \right\},$$

从而$\mathcal{N}(A) + \mathcal{R}(A) \neq \mathbb{F}^n$.

下面的定理给出了式(1.3.1)成立的充要条件.

定理 1.3.1 设$A \in \mathbb{F}^{n\times n}$, 则下列说法等价:

(i) $\mathcal{N}(A) \oplus \mathcal{R}(A) = \mathbb{F}^n$;

(ii) 存在可逆阵P和B使得$P^{-1}AP = \mathrm{diag}(0, B)$;

(iii) $\mathrm{rank}A = \mathrm{rank}A^2$;

(iv) $\mathcal{N}(A)\cap\mathcal{R}(A)=\{0\}$;

(v) $\mathcal{N}(A)+\mathcal{R}(A)=\mathbb{F}^n$.

证明 (i)⇒(ii). 取$\mathcal{N}(A)$的基x_1, $\cdots$, x_r, $\mathcal{R}(A)$的基x_{r+1}, $\cdots$, x_n. 由(i)和性质1.2.1知Ax_{r+1}, $\cdots$, Ax_n是$\mathcal{R}(A)$的基且x_1, $\cdots$, x_r, x_{r+1}, $\cdots$, x_n是$\mathbb{F}^n$的基, 于是存在可逆阵B使得

$$\begin{aligned}A\begin{bmatrix} x_{r+1} & \cdots & x_n\end{bmatrix} &= \begin{bmatrix} Ax_{r+1} & \cdots & Ax_n\end{bmatrix}\\ &= \begin{bmatrix} x_{r+1} & \cdots & x_n\end{bmatrix}B\\ &= \begin{bmatrix} x_1 & \cdots & x_r & x_{r+1} & \cdots & x_n\end{bmatrix}\begin{bmatrix} 0\\ B\end{bmatrix},\end{aligned}$$

故

$$A\begin{bmatrix} x_1 & \cdots & x_n\end{bmatrix}=\begin{bmatrix} x_1 & \cdots & x_n\end{bmatrix}\mathrm{diag}(0,B).$$

令$P=\begin{bmatrix} x_1 & \cdots & x_n\end{bmatrix}$, 则$P^{-1}AP=\mathrm{diag}(0,B)$.

(ii)⇒(iii). 此结论显然成立.

(iii)⇒(iv). 任取$w\in\mathcal{N}(A)\cap\mathcal{R}(A)$, 则$w\in\mathcal{N}(A)$且$w\in\mathcal{R}(A)$, 于是$Aw=0$, 且存在$x\in\mathbb{F}^n$使得$w=Ax$, 从而$A^2x=0$, 即$x\in\mathcal{N}(A^2)$. 这和$\mathcal{N}(A)\subseteq\mathcal{N}(A^2)$及

$$\dim\mathcal{N}(A)=n-\mathrm{rank}A=n-\mathrm{rank}A^2=\dim\mathcal{N}(A^2),$$

一起推出$x\in\mathcal{N}(A)$, 所以$w=Ax=0$. 故$\mathcal{N}(A)\cap\mathcal{R}(A)=\{0\}$.

(iv)⇒(v). 由(iv)及$\dim\mathcal{R}(A)+\dim\mathcal{N}(A)=n$推出

$$\dim(\mathcal{R}(A)+\mathcal{N}(A))=\dim\mathcal{R}(A)+\dim\mathcal{N}(A)-\dim(\mathcal{R}(A)\cap\mathcal{N}(A))=n,$$

于是$\mathcal{R}(A)+\mathcal{N}(A)=\mathbb{F}^n$.

(v)⇒(i). 由命题1.3.1得到. □

习 题

1. 若$A\in\mathbb{F}^{m\times n}$, $B\in\mathbb{F}^{p\times n}$, 则$\mathcal{N}\left(\begin{bmatrix} A\\ B\end{bmatrix}\right)=\mathcal{N}(A)\cap\mathcal{N}(B)$.

2. 若$A\in\mathbb{F}^{m\times n}$, $B\in\mathbb{F}^{n\times s}$, 则$\mathcal{R}(AB)=\mathcal{R}(A)$的充要条件是$\mathrm{rank}(AB)=\mathrm{rank}A$.

3. 若$A\in\mathbb{F}^{m\times n}$, $B\in\mathbb{F}^{n\times s}$, 则$\mathrm{rank}(AB)=\mathrm{rank}B$的充要条件是$\mathcal{N}(AB)=\mathcal{N}(B)$.

4. 设$S=\{Ax:Bx=0\}$, 其中$A\in\mathbb{F}^{m\times s}$, $B\in\mathbb{F}^{n\times s}$, 证明:

$$\dim S=\dim(\mathcal{N}(A)+\mathcal{N}(B))-\dim\mathcal{N}(A).$$

第2章　矩阵的分解与标准形

矩阵标准形在矩阵分析中有着非常重要的作用, 它起到简化问题的作用. 每种标准形都给处理某一类问题带来方便, 如等价标准形可用来处理许多与秩、行列式相关的问题; Fitting分解可用来处理与矩阵Drazin广义逆相关的问题; 奇异值分解可用来解决与矩阵Moore-Penrose广义逆相关的问题.

2.1　矩阵的等价分解

矩阵的等价分解是最常用的矩阵分解, 有些矩阵分解可基于等价分解给出证明, 例如满秩分解、Fitting分解等.

定理 2.1.1 (**等价分解**)　若$A \in \mathbb{F}^{m\times n}$, 则存在可逆阵$P$及$Q$, 使得

$$A = P\mathrm{diag}(I_r, 0)Q, \tag{2.1.1}$$

其中$r = \mathrm{rank}A$.

证明　设$x_{r+1}, \cdots, x_n$是$\mathcal{N}(A)$的基, 将其扩充成$\mathbb{F}^n$的基$x_1, \cdots, x_r, x_{r+1}, \cdots, x_n$. 使用性质1.2.1推出$Ax_1, \cdots, Ax_r$线性无关, 将其扩充成$\mathbb{F}^m$的基$Ax_1, \cdots, Ax_r, y_{r+1}, \cdots, y_m$. 令

$$P = \begin{bmatrix} Ax_1 & \cdots & Ax_r & y_{r+1} & \cdots & y_m \end{bmatrix},$$

及

$$Q = \begin{bmatrix} x_1 & \cdots & x_r & x_{r+1} & \cdots & x_n \end{bmatrix}^{-1},$$

则

$$AQ^{-1} = \begin{bmatrix} Ax_1 & \cdots & Ax_r & 0 & \cdots & 0 \end{bmatrix} = P\mathrm{diag}(I_r, 0),$$

于是$A = P\mathrm{diag}(I_r, 0)Q$. □

称$\mathrm{diag}(I_r, 0)$为矩阵A的等价标准形.

推论 2.1.1 若$A \in \mathbb{F}^{m\times n}$为行满秩, 则存在可逆阵$T$使得

$$A = \begin{bmatrix} I_m & 0 \end{bmatrix} T.$$

证明 由定理2.1.1知存在可逆阵P和Q满足

$$A = P \begin{bmatrix} I_r & 0 \end{bmatrix} Q,$$

其中$r = \mathrm{rank}A$. 令$T = \mathrm{diag}(P, I)Q$, 则T可逆且

$$A = \begin{bmatrix} P & 0 \end{bmatrix} Q = \begin{bmatrix} I_r & 0 \end{bmatrix} T.$$

□

推论 2.1.2 若$A \in \mathbb{F}^{m\times n}$为列满秩, 则存在可逆阵$S$使得$A = S \begin{bmatrix} I_n \\ 0 \end{bmatrix}$.

证明 类似于推论2.1.1. □

定理 2.1.2 (满秩分解) 若$A \in \mathbb{F}^{m\times n}$, 则存在列满秩矩阵$B \in \mathbb{F}^{m\times r}$和行满秩矩阵$C \in \mathbb{F}^{r\times n}$使得$A = BC$, 其中$r = \mathrm{rank}A$.

证明 由定理2.1.1知存在可逆阵P和Q满足$A = P\mathrm{diag}(I_r, 0)Q$, 其中$r = \mathrm{rank}A$. 令

$$B = P \begin{bmatrix} I_r \\ 0 \end{bmatrix}, \ C = \begin{bmatrix} I_r & 0 \end{bmatrix} Q,$$

则B为列满秩, C为行满秩, 并且$A = BC$. □

2.2 矩阵的Fitting分解

矩阵的Fitting分解将矩阵分解为可逆部分和幂零部分的直和, 可看做是复矩阵的Jordan分解(见2.5)在一般域上的推广. 矩阵的Fitting分解可用来解决许多与矩阵Drazin逆相关的问题(见9.2).

定义 2.2.1 设$N \in \mathbb{F}^{n\times n}$, 若存在正整数$k \geqslant 1$使得$N^k = 0$, 称$N$是幂零阵.

命题 2.2.1 若$N \in \mathbb{F}^{n\times n}$是幂零矩阵, 则

(i) N的特征值全为0;

(ii) $I_n \pm N$可逆.

证明 留给读者. □

命题 2.2.2 $A \in \mathbb{F}^{n\times n}$, $B \in \mathbb{F}^{n\times m}$, $M = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}$, 则$A$幂零的充要条件是$M$幂零.

证明 由

$$M^{k+1} = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}^{k+1} = \begin{bmatrix} A^{k+1} & A^kB \\ 0 & 0 \end{bmatrix}, \ \forall k \geqslant 1$$

易见. □

定理 2.2.1 (Fitting分解) 设$A \in \mathbb{F}^{n\times n}$, 则存在可逆阵$T$使得

$$A = T\mathrm{diag}(D, N)T^{-1},$$

其中D为可逆阵, N为幂零阵.

证明 对n用数学归纳法. 当$n = 1$时, 显然. 假设矩阵的阶数小于n时结论成立. 下面往证当矩阵的阶数是n时结论仍成立.

若$A = 0$或A可逆, 结论显然成立. 若$1 \leqslant \mathrm{rank}A = r < n$, 设$A$的等价分解为$A = P\mathrm{diag}(I_r, 0)Q$, 其中$P, Q$可逆. 令

$$QP = \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix}, \ Q_1 \in \mathbb{F}^{r\times r},$$

则

$$A = P\mathrm{diag}(I_r, 0)QPP^{-1} = P\begin{bmatrix} Q_1 & Q_2 \\ 0 & 0 \end{bmatrix}P^{-1}.$$

由归纳假设, 存在可逆阵P_1使得$Q_1 = P_1\mathrm{diag}(D, N_1)P_1^{-1}$, 其中$D$为可逆阵, N_1为幂零阵, 于是

$$\begin{aligned} A &= P\begin{bmatrix} P_1\mathrm{diag}(D, N_1)P_1^{-1} & Q_2 \\ 0 & 0 \end{bmatrix}P^{-1} \\ &= P\mathrm{diag}(P_1, I)\begin{bmatrix} D & 0 & Q_{21} \\ 0 & N_1 & Q_{22} \\ 0 & 0 & 0 \end{bmatrix}\mathrm{diag}(P_1, I)^{-1}P^{-1}, \end{aligned}$$

其中$\begin{bmatrix} Q_{21} \\ Q_{22} \end{bmatrix} = P_1^{-1}Q_2$. 令

$$T = P\mathrm{diag}(P_1, I)\begin{bmatrix} I & 0 & -D^{-1}Q_{21} \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}, \quad N = \begin{bmatrix} N_1 & Q_{22} \\ 0 & 0 \end{bmatrix},$$

则$A = T\mathrm{diag}(D, N)T^{-1}$. 现在只需证$N$为幂零阵. 事实上, 由$N_1$为幂零阵及命题2.2.2知$N$为幂零阵. □

2.3 复(实)矩阵的奇异值分解

由于矩阵奇异值分解有好的数值稳定性, 它已被用于设计求解许多问题的数值方法. 因而矩阵奇异值分解已被应用到许多领域.

定义 2.3.1 若$A \in \mathbb{R}^{n\times n}$满足$AA^T = A^TA = I_n$, 称$A$为正交阵. 若$A \in \mathbb{C}^{n\times n}$满足$AA^* = A^*A = I_n$, 称$A$为酉矩阵.

定理 2.3.1 (实矩阵的奇异值分解) 设$A \in \mathbb{R}^{m\times n}$, 则存在正交阵$U$, V使得

$$A = U\mathrm{diag}(\Sigma, 0)V,$$

其中$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$, $\sigma_i > 0$, $i = 1, 2, \cdots, r$.

定理 2.3.2 (复矩阵的奇异值分解) 设$A \in \mathbb{C}^{m\times n}$, 则存在酉矩阵$U$, V使得

$$A = U\mathrm{diag}(\Sigma, 0)V,$$

其中$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$, $\sigma_i > 0$, $i = 1, 2, \cdots, r$.

上面两个定理的证明见参考文献[2]. 易见上面两个定理中的σ_1, σ_2, $\cdots$, σ_r为矩阵A的非零奇异值. 显然σ_i^2, $i = 1, 2, \cdots, r$是A^*A(或AA^*)的所有非零特征值.

令$P = U\mathrm{diag}(\Sigma, I_{m-r})$及$Q = \mathrm{diag}(\Sigma, I_{n-r})V$, 则

$$A = P\mathrm{diag}(I_r, 0)V = U\mathrm{diag}(I_r, 0)Q$$

为矩阵A的等价分解. 这提供了使用MATLAB工具软件的svd函数计算矩阵等价分解的方法.

2.4 矩阵的对角化

对于矩阵$A \in \mathbb{F}^{n\times n}$, 若存在可逆阵$P$使得$P^{-1}AP$为对角矩阵, 则称矩阵$A$可对角化. 关于矩阵的对角化问题, 有下面的结论.

定理 2.4.1 [2] 设$A \in \mathbb{F}^{n\times n}$, 则$A$可对角化的充要条件是$A$有$n$个线性无关的特征向量.

尽管定理2.4.1给出了矩阵可对角化的充要条件, 但事实上很多矩阵不是可对角化的. 下面讨论复Hermite矩阵的对角化问题.

定义 2.4.1 若$A \in \mathbb{C}^{n\times n}$满足$AA^* = A^*A$, 称$A$是正规矩阵.

命题 2.4.1 上(下)三角的正规矩阵一定是对角阵.

证明 留给读者. □

定理 2.4.2 设$A \in \mathbb{C}^{n\times n}$, 则存在酉矩阵$U$使得

$$U^*AU = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ \star & & \lambda_n \end{bmatrix},$$

其中$\lambda_1, \cdots, \lambda_n$是矩阵$A$的特征值.

证明 对n应用数学归纳法. 当$n = 1$时结论显然成立. 假设$n = k-1$时结论成立. 当$n = k$时, 取A的一个特征值λ和其对应的一个单位特征向量u, 则

$$Au = \lambda u.$$

将u扩充成酉空间$\mathbb{C}^n$的一个标准正交基$u_1, \cdots, u_{k-1}, u$, 于是

$$A\begin{bmatrix} u_1 & \cdots & u_{k-1} & u \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_{k-1} & u \end{bmatrix}\begin{bmatrix} A_1 & 0 \\ x^* & \lambda \end{bmatrix}, \qquad (2.4.1)$$

其中$A_1 \in \mathbb{C}^{(k-1)\times(k-1)}$, $x \in \mathbb{C}^{k-1}$. 由归纳假设, 存在酉矩阵U_1使得

$$U_1^*A_1U_1 = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ \star & & \lambda_{k-1} \end{bmatrix}, \qquad (2.4.2)$$

其中$\lambda_1, \cdots, \lambda_{k-1}$是矩阵$A_1$的特征值. 令

$$U = \begin{bmatrix} u_1 & \cdots & u_{k-1} & u \end{bmatrix}\mathrm{diag}(U_1, 1).$$

组合式(2.4.1)和(2.4.2)得

$$U^*AU = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_{k-1} & \\ \star & & & \lambda \end{bmatrix},$$

即$n = k$时结论成立. □

使用定理2.4.2容易得到:

定理 2.4.3 正规矩阵一定酉相似于对角阵.

推论 2.4.1 设$A^* = A \in \mathbb{C}^{n\times n}$, 则

(i) A的特征值一定是实数;

(ii) 存在酉矩阵U使得

$$A = U\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)U^*, \tag{2.4.3}$$

其中λ_1, λ_2, $\cdots$, λ_n是矩阵A的特征值.

若将式(2.4.3)中的U分块为$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$, 则

$$A = \lambda_1 u_1 u_1^* + \lambda_2 u_2 u_2^* + \cdots + \lambda_n u_n u_n^*. \tag{2.4.4}$$

称式(2.4.4)为Hermite矩阵A的谱分解.

定理 2.4.4 设$A^* = A \in \mathbb{C}^{n\times n}$, 则

$$\max_{0\neq x\in\mathbb{C}^n} \frac{x^*Ax}{x^*x} = \lambda_{\max}(A), \quad \min_{0\neq x\in\mathbb{C}^n} \frac{x^*Ax}{x^*x} = \lambda_{\min}(A).$$

证明 设$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$是$A$的特征值, 则存在酉矩阵$Q$使得

$$A = Q\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)Q^*,$$

故

$$\frac{x^*Ax}{x^*x} = \frac{x^*Q\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)Q^*x}{x^*QQ^*x}.$$

令

$$y = Q^*x = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T,$$

则

$$\frac{x^*Ax}{x^*x} = \frac{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2}{y_1^2 + y_2^2 + \cdots + y_n^2}.$$

因

$$\lambda_1(y_1^2 + y_2^2 + \cdots + y_n^2) \geqslant (\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2) \geqslant \lambda_n(y_1^2 + y_2^2 + \cdots + y_n^2),$$

所以$\lambda_1 \geqslant \frac{x^*Ax}{x^*x} \geqslant \lambda_n$, $\quad \forall\ 0 \neq x \in \mathbb{C}^n$. 取$x_1$是属于$\lambda_1$的特征向量, x_n是属于λ_n的特征向量, 则

$$\lambda_1 = \frac{x_1^*Ax_1}{x_1^*x_1}, \quad \lambda_n = \frac{x_n^*Ax_n}{x_n^*x_n},$$

因此结论成立. □

2.5 复矩阵的Jordan分解

上节证明了复Hermite矩阵一定是可对角化的, 这个结论对于一般的复方阵不成立(请读者举例子), 然而下面的定理表明每个复方阵都可相似于一个Jordan形阵.

定理 2.5.1 (Jordan**分解**) 设$A \in \mathbb{C}^{n\times n}$, 则存在可逆阵$P$使得

$$A = P\text{diag}(J_1, J_2, \cdots, J_t)P^{-1}, \tag{2.5.1}$$

其中

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & 1 \\ & & \lambda_i \end{bmatrix}, \ i = 1, 2, \cdots, t.$$

定理2.5.1的证明见参考文献[3]. 称式(2.5.1)中$\text{diag}(J_1, J_2, \cdots, J_t)$为复矩阵$A$的Jordan标准形, 称$J_1$, J_2, $\cdots$, J_t为复矩阵A的Jordan块.

注记 2.5.1 Jordan块的写法不是唯一的, 例如上面的J_1也可写成

$$J_1 = \begin{bmatrix} \lambda_1 & a & & \\ & \ddots & \ddots & \\ & & \ddots & a \\ & & & \lambda_1 \end{bmatrix}, \ a \neq 0$$

或

$$J_1 = \begin{bmatrix} \lambda_1 & & & \\ b & \ddots & & \\ & \ddots & \ddots & \\ & & b & \lambda_1 \end{bmatrix}, \ b \neq 0.$$

最后介绍下面的定理, 证明可参考相关文献.

定理 2.5.2 设矩阵$E\in\mathbb{C}^{n\times n}$的Jordan标准形为$E=TJT^{-1}$, 并设

$$\begin{cases} J=\operatorname{diag}(J_1,\ \cdots,\ J_l), \\ J_a=\operatorname{diag}(J_{a1},\ \cdots,\ J_{aq_a}),\ a=1,2,\cdots,l, \\ J_{ab}=s_aI+N_{p_{ab}},\ b=1,2,\cdots,q_a,\ a=1,2,\cdots,l, \end{cases}$$

$$\begin{cases} T=\begin{bmatrix} T_1 & \cdots & T_l \end{bmatrix}, \\ T_a=\begin{bmatrix} T_{a1} & \cdots & T_{aq_a} \end{bmatrix},\ a=1,2,\cdots,l, \\ T_{ab}=\begin{bmatrix} t_{ab}^1 & \cdots & t_{ab}^{p_{ab}} \end{bmatrix},\ b=1,2,\cdots,q_a,\ a=1,2,\cdots,l, \end{cases}$$

其中$s_1,\cdots,s_l$彼此互异, $t_{ab}^c\in\mathbb{C}^n$, $p_{a1}\geqslant p_{a2}\geqslant\cdots\geqslant p_{aq_a}$; $a=1,2,\cdots,l$,

$$N_k=\begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}\in\mathbb{R}^{k\times k},$$

则

(i) $\mathcal{N}\left((s_aI-E)^{p_{a1}}\right)=\operatorname{span}\left\{t_{ab}^k:b=1,2,\cdots,q_a,\ k=1,2,\cdots,p_{ab}\right\}$, $a=1,2,\cdots,l$; (见参考文献[4])

(ii) $\mathbb{C}^n=\bigoplus\limits_{a=1}^{l}\mathcal{N}\left((s_aI-E)^{p_{a1}}\right)$. (见参考文献[5]162页)

2.6 实对称矩阵的惯性指数分解

前面讨论的矩阵奇异值分解中含有矩阵的所有奇异值, 复Hermite阵酉相似的对角阵中含有矩阵的特征值. 然而, 有时我们只关心一个实对称矩阵的正、负和零特征值个数(如控制理论中的稳定性判据), 而不关心特征值的具体值. 本节讨论的惯性指数分解就是含有正、负和零特征值的个数的分解.

定理 2.6.1 (惯性指数分解) 设$A^T=A\in\mathbb{R}^{n\times n}$, 则存在可逆阵$P$使得$A=P\operatorname{diag}(I_s,-I_t,0)P^T$. 称$s$为$A$的正惯性指数, t为A的负惯性指数.

推论 2.6.1 若实矩阵$A\geqslant 0$, 则

(i) 存在可逆阵P使得$A = P\text{diag}(I_s, 0)P^T$, 其中$s$是$A$的正惯性指数;

(ii) 存在列满秩矩阵B使得$A = BB^T$.

习 题

1. 设$A \in \mathbb{F}^{n\times n}$, 若存在正整数$k \geqslant 2$使得$A^k = A$, 称$A$为$k-$Potent阵. 证明: 若$A$为$k$-Potent阵, 则存在可逆阵$P$使得$A = P\text{diag}(D, 0)P^{-1}$, 其中$D^{k-1} = I$.

2. 证明: 若A是幂等阵(即2$-$Potent阵), 则存在可逆阵P使得

$$A = P\text{diag}(I_r, 0)P^{-1},$$

   其中$r = \text{rank}A$.

3. 设$A \in \mathbb{F}^{n\times n}$满足$A^2 = I_n$(称$A$为对合阵), 证明: 存在可逆阵$P$使得$A = P\text{diag}(I_r, -I_{n-r})P^{-1}$, 其中$r = \text{rank}A$.

4. 设$A \in \mathbb{F}^{m\times n}$有分解式$A = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix}$, 其中$X \in \mathbb{F}^{r\times r}$, $\text{rank}X = \text{rank}A = r$, 证明: A有如下形式的满秩分解

$$A = \begin{bmatrix} X \\ Z \end{bmatrix} [I_r \ \ X^{-1}Y] = \begin{bmatrix} I_r \\ ZX^{-1} \end{bmatrix} [X \ \ Y].$$

5. 设$A, B \in \mathbb{F}^{n\times n}$满足$AB = BA$且$A$奇异, 并设$A$的Fitting分解为$A = P\text{diag}(D, N)P^{-1}$, 则$B = P\text{diag}(B_1, B_4)P^{-1}$, 其中$B_1$和$B_4$有适当的维数, 并且满足$DB_1 = B_1D$且$NB_4 = B_4N$.

6. 已知$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$, 求$A$的奇异值分解.

7. 设A的奇异值分解为$A = U\text{diag}(\Sigma, 0)V^*$, $\Sigma = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$, 记$U$和$V$的列分别为$u_1$, u_2, $\cdots$, u_m和v_1, v_2, $\cdots$, v_n, 证明:

   (i) $\mathcal{N}(A) = \text{span}\{v_{r+1}, v_{r+2}, \cdots, v_n\}$;

   (ii) $\mathcal{R}(A) = \text{span}\{u_1, u_2, \cdots, u_r\}$;

   (iii) $A = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots + \sigma_r u_r v_r^*$.

8. 设$M \in \mathbb{C}^{m\times n}$, $x \in \mathbb{C}$, 则$\sigma_{\max}(xM) = |x|\sigma_{\max}(M)$.

9. 设$A \in \mathbb{C}^{m\times n}$, 证明:

$$\mathcal{N}(A) = (\mathcal{R}(A^*))^{\perp},\ \mathcal{R}(A) = (\mathcal{N}(A^*))^{\perp},$$

其中上标"⊥"表示子空间的正交补. 特别地, 若A是Hermite阵(即$A^* = A$), 则

$$\mathcal{N}(A) = (\mathcal{R}(A))^{\perp},\ \mathcal{R}(A) = (\mathcal{N}(A))^{\perp}.$$

10. 若$A \in \mathbb{C}^{m\times n}$和$B \in \mathbb{C}^{m\times s}$满足$\mathcal{R}(A) \cap \mathcal{R}(B) = \{0\}$, 则$\dim \mathcal{R}(A^* B_{\perp}) = \dim \mathcal{R}(A^*)$, 其中$B_{\perp}$是使得$\mathcal{R}(B_{\perp}) = \mathcal{R}(B)^{\perp}$的任意矩阵. (提示: 使用关系$\dim \mathcal{R}(XY) = \dim \mathcal{R}(Y) - \dim(\mathcal{R}(Y) \cap \mathcal{N}(X))$, 并利用上题.)

11. 设$A \in \mathbb{C}^{m\times n}$, 证明: $\mathrm{rank}(A^*A) = \mathrm{rank}(AA^*) = \mathrm{rank}(A)$.

12. 设正规矩阵A的特征值互异, 证明: 矩阵方程$AX = XA$的任意解均为正规矩阵.

13. 设A是实对称幂等阵, 则存在正交阵Q使得$Q^TAQ = \mathrm{diag}(I_r, 0)$, 其中$r = \mathrm{rank}A$.

14. 若$\begin{bmatrix} A_1 & A_2 \\ A_2^T & 0 \end{bmatrix}$是实对称半正定阵, 则$A_2 = 0$.

15. 若$A \geqslant 0$, 则存在唯一的$B \geqslant 0$使得$A = B^2$.

16. 设$A = \begin{bmatrix} 16 & 8 \\ 8 & 4 \end{bmatrix}$, 求$A^{\frac{1}{2}}$.

17. 设$H = [h_{ij}] \in \mathbb{R}^{n\times n}$满足$H \geqslant 0$, 证明: $\det H \leqslant h_{11}h_{22}\cdots h_{nn}$.

第3章　矩阵对的分解和标准形

上一章介绍了几种常用的矩阵的分解及标准形, 但有时为了解决一些应用领域中提出的问题, 需要同时考虑两个矩阵或更多个矩阵的分解和标准形. 本节介绍几种矩阵对的分解及标准形, 这些结果有的可用于解决控制理论中的问题[6, 7], 有的可用于解决线性(加法)保持问题[8].

3.1　(非)正则矩阵对的等价标准形

本节研究矩阵对的一种等价标准形(即下面的定理), 这是研究广义线性系统理论的基础[9].

定理 3.1.1　设$M,N\in\mathbb{R}^{m\times n}$, 则存在可逆阵$P\in\mathbb{R}^{m\times m}$和$Q\in\mathbb{R}^{n\times n}$使得

$$\begin{aligned}P(sM-N)Q=&\operatorname{diag}(0_{m_0\times n_0},L_{u_1}(s),\cdots,L_{u_p}(s),(L_{v_1}(s))^T,\cdots,\\&(L_{v_q}(s))^T,sI_{w_1}-N_1,sM_2-I_{w_2}),\ \forall s\in\mathbb{C},\end{aligned}\tag{3.1.1}$$

其中$M_2\in\mathbb{R}^{w_2\times w_2}$, $N_1\in\mathbb{R}^{w_1\times w_1}$, M_2是幂零矩阵,

$$L_k(s)=\begin{bmatrix}s&-1&0&\cdots&0\\0&s&-1&\cdots&0\\\vdots&\vdots&\ddots&\ddots&\vdots\\0&0&\cdots&s&-1\end{bmatrix}\in\mathbb{R}^{k\times(k+1)},$$

$$m_0+\sum_{i=1}^{p}u_i+\sum_{j=1}^{q}v_j+q+w_1+w_2=m,$$

$$n_0+\sum_{i=1}^{p}u_i+\sum_{j=1}^{q}v_j+p+w_1+w_2=n.$$

定理3.1.1的证明见附录A.

通常称式(3.1.1)的右端为矩阵对(M,N)的Kronecker标准形.

定义 3.1.1 设$E,A\in\mathbb{R}^{n\times n}$, 若存在$s\in\mathbb{C}$满足

$$\det(sE-A)\neq 0, \tag{3.1.2}$$

称矩阵对(E,A)是正则的.

显然, 当矩阵对(E,A)正则时, 不满足式(3.1.2)的复数s至多为n个.

定理 3.1.2 设$M,N\in\mathbb{R}^{n\times n}$, 则矩阵对$(M,N)$正则的充要条件是存在可逆阵$P$, $Q\in\mathbb{R}^{n\times n}$使得

$$P(sM-N)Q=\mathrm{diag}(sI_{n_1}-N_1,sM_2-I_{n_2}),\ \forall s\in\mathbb{C}, \tag{3.1.3}$$

其中$N_1\in\mathbb{R}^{n_1\times n_1}$, $M_2\in\mathbb{R}^{n_2\times n_2}$, $n_1+n_2=n$, M_2为幂零矩阵.

称式(3.1.3)的右端为矩阵对(M,N)的Weierstrass标准形.

证明 必要性由命题A.0.4得到, 下证充分性.

由式(3.1.3)得

$$\det(sM-N)=\det(PQ)^{-1}\det(sI-N_1)\det(sM_2-I).$$

由M_2为幂零矩阵知

$$\det(sM_2-I)=(-1)^{n_2}\neq 0,\ \forall s\in\mathbb{C}.$$

而当s不是N_1的特征根时$\det(sI-N_1)\neq 0$. 故存在$s_0\in\mathbb{C}$使$\det(s_0M-N)\neq 0$, 即(M,N)正则. □

3.2 矩阵对的能控能观结构分解

本节介绍矩阵对的能控能观结构分解. 矩阵对的能控能观结构分解在线性系统理论中有重要应用.

对于矩阵$X\in\mathbb{R}^{n\times n}$, $Y\in\mathbb{R}^{n\times m}$和正整数$k$, 用$C_{k,X,Y}$表示矩阵

$$\begin{bmatrix} Y & XY & X^2Y & \cdots & X^{k-1}Y \end{bmatrix}.$$

定义 3.2.1 若$A\in\mathbb{R}^{n\times n}$和$B\in\mathbb{R}^{n\times r}$满足$\mathrm{rank}C_{n,A,B}=n$, 则称矩阵对$(A,B)$能控. 否则称矩阵对$(A,B)$不能控.

定义 3.2.2 若$A \in \mathbb{R}^{n\times n}$和$C \in \mathbb{R}^{m\times n}$满足$\text{rank}C_{n,A^T,C^T} = n$, 则称矩阵对$(A, C)$能观; 否则称矩阵对$(A, C)$不能观.

从上面的两个定义易见

$$(A, B)\text{能控} \Leftrightarrow (A^T, B^T)\text{能观}. \tag{3.2.1}$$

定理 3.2.1 设$A \in \mathbb{R}^{n\times n}$和$B \in \mathbb{R}^{n\times r}$满足$(A, B)$不能控, 则存在可逆阵$P \in \mathbb{R}^{n\times n}$满足

$$P^{-1}AP = \begin{bmatrix} A_c & A_{12} \\ 0 & A_{\bar{c}} \end{bmatrix}, \quad P^{-1}B = \begin{bmatrix} B_c \\ 0 \end{bmatrix}, \tag{3.2.2}$$

且满足(A_c, B_c)能控、$(A_{\bar{c}}, 0)$不能控, 其中$B_c \in \mathbb{R}^{n_c\times r}$, $A_c \in \mathbb{R}^{n_c\times n_c}$, $A_{\bar{c}} \in \mathbb{R}^{n_{\bar{c}}\times n_{\bar{c}}}$, $A_{12} \in \mathbb{R}^{n_c\times n_{\bar{c}}}$, $n_c + n_{\bar{c}} = n$.

证明 由(A, B)不能控知$\text{rank}C_{n,A,B} < n$. 设$p_1, \cdots, p_{n_c}$是$\mathcal{R}(C_{n,A,B})$的基, 将其扩充成$\mathbb{R}^n$的基$p_1, \cdots, p_{n_c}, p_{n_c+1}, \cdots, p_n$. 令

$$P = \begin{bmatrix} p_1 & \cdots & p_{n_c} & p_{n_c+1} & \cdots & p_n \end{bmatrix}.$$

因$Ap_1, \cdots, Ap_{n_c}$的各列为$\mathcal{R}(C_{n,A,B})$中的向量, 故B的各列和$Ap_1, \cdots, Ap_{n_c}$可由$p_1, \cdots, p_{n_c}$线性表出, 于是

$$B = P\begin{bmatrix} B_c \\ 0 \end{bmatrix},$$

且

$$AP = P\begin{bmatrix} A_c & A_{12} \\ 0 & A_{\bar{c}} \end{bmatrix},$$

其中$B_c \in \mathbb{R}^{n_c\times r}$, $A_c \in \mathbb{R}^{n_c\times n_c}$, $A_{\bar{c}} \in \mathbb{R}^{n_{\bar{c}}\times n_{\bar{c}}}$, $A_{12} \in \mathbb{R}^{n_c\times n_{\bar{c}}}$, $n_c + n_{\bar{c}} = n$. 因而式(3.2.2)成立, 且

$$\begin{aligned} C_{n,A,B} &= \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} \\ &= P\begin{bmatrix} B_c & A_cB_c & \cdots & A_c^{n-1}B_c \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \end{aligned}$$

由Hamilton-Cayley定理得

$$\begin{aligned} & \text{rank}\begin{bmatrix} B_c & A_cB_c & \cdots & A_c^{n_c-1}B_c \end{bmatrix} \\ = & \text{rank}\begin{bmatrix} B_c & A_cB_c & \cdots & A_c^{n-1}B_c \end{bmatrix} \\ = & \text{rank}\begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix} \\ = & \text{rank}C_{n,A,B} \\ = & n_c, \end{aligned}$$

即(A_c, B_c)能控. 另一方面, 显然有$(A_{\bar{c}}, 0)$不能控. □

由式(3.2.1)及定理3.2.1容易推出下面的定理.

定理 3.2.2 设$A \in \mathbb{R}^{n\times n}$和$C \in \mathbb{R}^{m\times n}$满足$(A, C)$不能观, 则存在可逆阵$P \in \mathbb{R}^{n\times n}$满足

$$P^{-1}AP = \begin{bmatrix} A_o & 0 \\ A_{21} & A_{\bar{o}} \end{bmatrix}, \quad CP = \begin{bmatrix} C_o & 0 \end{bmatrix}, \tag{3.2.3}$$

且满足(A_o, C_o)能观、$(A_{\bar{o}}, 0)$不能观, 其中$C_o \in \mathbb{R}^{m\times n_o}$, $A_o \in \mathbb{R}^{n_o\times n_o}$, $A_{\bar{o}} \in \mathbb{R}^{n_{\bar{o}}\times n_{\bar{o}}}$, $A_{21} \in \mathbb{R}^{n_{\bar{o}}\times n_o}$, $n_o + n_{\bar{o}} = n$.

称式(3.2.2)为不能控矩阵对(A, B)的能控性分解, 式(3.2.3)为不能观矩阵对(A, C)的能观性分解. 下面进一步给出(A, B, C)的结构分解.

定理 3.2.3 (同时按能控和能观分解) 设$A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$和$C \in \mathbb{R}^{q\times n}$满足$(A, C)$不能观且$(A, B)$不能控, 则存在可逆阵$P \in \mathbb{R}^{n\times n}$满足

$$P^{-1}AP = \begin{bmatrix} A_{11} & 0 & A_{13} & 0 \\ A_{21} & A_{22} & A_{23} & A_{24} \\ 0 & 0 & A_{33} & 0 \\ 0 & 0 & A_{43} & A_{44} \end{bmatrix}, \ P^{-1}B = \begin{bmatrix} B_{11} \\ B_{21} \\ 0 \\ 0 \end{bmatrix},$$

$$CP = \begin{bmatrix} C_{11} & 0 & C_{13} & 0 \end{bmatrix}, \tag{3.2.4}$$

其中

(i) (A_{11}, B_{11})能控且(A_{11}, C_{11})能观;

(ii) $\left(\begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}, \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}\right)$能控, $\left(\begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}, \begin{bmatrix} C_{11} & 0 \end{bmatrix}\right)$不能观;

(iii) $\left(\begin{bmatrix} A_{11} & A_{13} \\ 0 & A_{33} \end{bmatrix}, \begin{bmatrix} B_{11} \\ 0 \end{bmatrix}\right)$不能控, $\left(\begin{bmatrix} A_{11} & A_{13} \\ 0 & A_{33} \end{bmatrix}, \begin{bmatrix} C_{11} & C_{13} \end{bmatrix}\right)$能观;

(iv) $(A_{44}, 0)$不能控且不能观.

证明 设$\xi_1,\cdots,\xi_k$是

$$\mathcal{R}(C_{n,A,B})\cap\mathcal{N}(C^T_{n,A^T,C^T})$$

的基, 将其扩充为$\mathcal{R}(C_{n,A,B})$的基

$$\xi_1,\xi_2,\cdots,\xi_k,\eta_1,\cdots,\eta_p$$

及$\mathcal{N}(C^T_{n,A^T,C^T})$的基

$$\xi_1,\xi_2,\cdots,\xi_k,\varepsilon_1,\cdots,\varepsilon_q.$$

易见向量组

$$\xi_1,\xi_2,\cdots,\xi_k,\eta_1,\cdots,\eta_p,\varepsilon_1,\cdots,\varepsilon_q$$

线性无关, 再将其扩充为$\mathbb{R}^n$的基

$$\xi_1,\xi_2,\cdots,\xi_k,\eta_1,\cdots,\eta_p,\varepsilon_1,\cdots,\varepsilon_q,\delta_1,\cdots,\delta_l.$$

令

$$P=\begin{bmatrix}\eta_1 & \cdots & \eta_p & \xi_1 & \cdots & \xi_k & \delta_1 & \cdots & \delta_l & \varepsilon_1 & \cdots & \varepsilon_q\end{bmatrix},$$

使用$\mathcal{R}(C_{n,A,B})$和$\mathcal{N}(C^T_{n,A^T,C^T})$均是$A$的不变子空间容易推出式(3.2.4)成立.

显然(iv)成立, 下面证明(i)~(iii)成立. 记

$$A_1=\begin{bmatrix}A_{11} & 0\\ A_{21} & A_{22}\end{bmatrix},B_1=\begin{bmatrix}B_{11}\\ B_{21}\end{bmatrix},C_1=\begin{bmatrix}C_{11} & 0\end{bmatrix},$$

类似于定理3.2.1的证明可得

$$\mathrm{rank}C_{k+p,A_1,B_1}=\mathrm{rank}C_{n,A_1,B_1}=\mathrm{rank}C_{n,A,B}=k+p,$$

从而(A_1,B_1)能稳; 另一方面, 对于A_{22}的特征根λ_0有

$$\mathrm{rank}\begin{bmatrix}\lambda_0 I-A_1\\ C_1\end{bmatrix}<k+p,$$

于是(A_1,C_1)不能观, 故(ii)成立. 类似地, 可证得(iii)成立. 由(ii)和(iii)容易推出(i)成立. □

对于线性系统

$$\dot{x}(t)=Ax(t)+Bu(t), \tag{3.2.5a}$$

$$y(t)=Cx(t)+Du(t), \tag{3.2.5b}$$

记其传递函数为$C(sI-A)^{-1}B+D$为$G(s)$.

由定理3.2.3及线性系统的最小实现理论容易推出下面的定理.

定理 3.2.4 系统(3.2.5)满足

$$\lambda(A)=\lambda(G(s))\cup\lambda_{\overline{c}}(A,B)\cup\lambda_{\overline{o}}(A,C),$$

其中$\lambda(A)$表示A的所有特征根(称为系统(3.2.5)的极点)的集合, $\lambda_{\overline{c}}(A,B)$表示矩阵对$(A,B)$的所有不能控特征根的集合(即满足$\operatorname{rank}\begin{bmatrix}\lambda I_n-A & B\end{bmatrix}<n$的所有复数$\lambda$的集合), $\lambda_{\overline{o}}(A,C)$表示矩阵对$(A,C)$的所有不能观特征根的集合(即满足$\operatorname{rank}\begin{bmatrix}\lambda I_n-A\\ C\end{bmatrix}<n$的所有复数$\lambda$的集合), $\lambda(G(s))$表示传递函数矩阵$G(s)$的所有极点的集合(即$G(s)$的所有元素的最小公分母的零点).

3.3 能控矩阵对的规范形

本节介绍能控矩阵对的规范形, 并由此给出一些与线性系统极点配置[7]相关的结论.

定理 3.3.1 若$A\in\mathbb{R}^{n\times n}$和$B\in\mathbb{R}^{n\times m}$满足$(A,B)$能控且$B$列满秩, 则存在$P\in GL_n(\mathbb{R})$和正整数$n_1$, n_2, $\cdots$, n_m满足

$$P^{-1}AP=\begin{bmatrix}A_{11} & A_{12} & \cdots & A_{1m}\\ A_{21} & A_{22} & \cdots & A_{2m}\\ \cdots & \cdots & \cdots & \cdots\\ A_{m1} & A_{m2} & \cdots & A_{mm}\end{bmatrix},\tag{3.3.1}$$

$$P^{-1}B=\begin{bmatrix}B_1\\ B_2\\ \vdots\\ B_m\end{bmatrix},\tag{3.3.2}$$

其中$B_i\in\mathbb{R}^{n_i\times m}$和$A_{ij}\in\mathbb{R}^{n_i\times n_j}$, $i,j=1,2,\cdots,m$有下面的形式:

$$B_i=\begin{bmatrix}0 & 0 & 0\\ 0_{1\times(i-1)} & 1 & *\end{bmatrix},\tag{3.3.3}$$

$$A_{ii}=\begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ \vdots & & \ddots & \ddots & 0\\ 0 & \cdots & \cdots & 0 & 1\\ a_{ii}^{(0)} & a_{ii}^{(1)} & \cdots & a_{ii}^{(n_i-2)} & a_{ii}^{(n_i-1)}\end{bmatrix}, \tag{3.3.4}$$

$$A_{ij}=\begin{bmatrix}0 & 0 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & 0\\ a_{ij}^{(0)} & a_{ij}^{(1)} & \cdots & a_{ij}^{(n_j-1)}\end{bmatrix},\ i\neq j. \tag{3.3.5}$$

定理3.3.1的证明见附录B, 下面介绍定理3.3.1的应用.

推论 3.3.1 若$A\in\mathbb{R}^{n\times n}$和非零的$b\in\mathbb{R}^n$满足$(A,b)$能控, 则存在唯一的矩阵$P\in GL_n(\mathbb{R})$满足

$$P^{-1}AP=\begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ \vdots & & \ddots & \ddots & 0\\ 0 & \cdots & \cdots & 0 & 1\\ a_0 & a_1 & \cdots & a_{n-2} & a_{n-1}\end{bmatrix},\ P^{-1}b=\begin{bmatrix}0\\ \vdots\\ 0\\ 1\end{bmatrix}, \tag{3.3.6}$$

其中$a_i\in\mathbb{R},\ i=0,1,2,\cdots,n-1$.

证明 由定理3.3.1知存在可逆阵P满足式(3.3.6). 下证P的唯一性.

若存在可逆阵P_1和P_2满足式(3.3.6), 则

$$P_1\hat{A}P_1^{-1}=P_2\hat{A}P_2^{-1},\ P_1e_n=P_2e_n,$$

其中

$$\hat{A}=\begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ \vdots & & \ddots & \ddots & 0\\ 0 & \cdots & \cdots & 0 & 1\\ a_0 & a_1 & \cdots & a_{n-2} & a_{n-1}\end{bmatrix},\ e_n=\begin{bmatrix}0\\ \vdots\\ 0\\ 1\end{bmatrix}.$$

令$T=P_1^{-1}P_2$, 则

$$T\hat{A}=\hat{A}T \tag{3.3.7}$$

且

$$Te_n = e_n. \tag{3.3.8}$$

由式(3.3.8)易见

$$T = \begin{bmatrix} t_{11} & \cdots & t_{1,n-1} & 0 \\ \cdots & \cdots & \cdots & \cdots \\ t_{n-1,1} & \cdots & t_{n-1,n-1} & 0 \\ t_{n1} & \cdots & t_{n,n-1} & 1 \end{bmatrix}.$$

将T代入式(3.3.7)并比较两边对应位置元素易得$T = I_n$, 即$P_1 = P_2$. □

推论 3.3.2 若$A \in \mathbb{R}^{n\times n}$和$B \in \mathbb{R}^{n\times m}$满足$(A, B)$能控及$B$列满秩, 则存在$P \in GL_n(\mathbb{R})$, $Q \in GL_m(\mathbb{R})$和正整数$n_1, n_2, \cdots, n_m$满足

$$P^{-1}AP = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{bmatrix}, \tag{3.3.9}$$

$$P^{-1}BQ = \begin{bmatrix} B_{11} & 0 & \cdots & 0 \\ 0 & B_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & B_{mm} \end{bmatrix}, \tag{3.3.10}$$

其中$B_{ii} \in \mathbb{R}^{n_i}$和$A_{ij} \in \mathbb{R}^{n_i\times n_j}$, $i, j = 1, 2, \cdots, m$有下面的形式:

$$B_{ii} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}^T, \tag{3.3.11}$$

$$A_{ii} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ a_{ii}^{(0)} & a_{ii}^{(1)} & \cdots & a_{ii}^{(n_i-2)} & a_{ii}^{(n_i-1)} \end{bmatrix}, \tag{3.3.12}$$

$$A_{ij} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \\ a_{ij}^{(0)} & a_{ij}^{(1)} & \cdots & a_{ij}^{(n_j-1)} \end{bmatrix}, \ i \neq j. \tag{3.3.13}$$

证明 由定理3.3.1易见. □

推论 3.3.3 若$A \in \mathbb{R}^{n\times n}$和$B \in \mathbb{R}^{n\times m}$满足$(A,B)$能控, β_0, β_1, $\cdots$, $\beta_{n-1} \in \mathbb{R}$, 则存在$K \in \mathbb{R}^{m\times n}$使得

$$\det(sI - A - BK) = \beta_0 + \beta_1 s + \cdots + \beta_{n-1}s^{n-1} + s^n.$$

证明 不妨设B列满秩. 由(A,B)能控和推论3.3.2, 存在$P \in GL_n(\mathbb{R})$, $Q \in GL_m(\mathbb{R})$和正整数n_1, n_2, $\cdots$, n_m使得式(3.3.9)至式(3.3.13)成立. 令

$$K = Q\begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1m} \\ K_{21} & K_{22} & \cdots & K_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ K_{m1} & K_{m2} & \cdots & K_{mm} \end{bmatrix}P^{-1},$$

则由式(3.3.9)和(3.3.10)推出

$$\begin{aligned} & sI - A - BK \\ = \; & P\begin{bmatrix} sI - A_{11} - B_{11}K_{11} & \cdots & -A_{1m} - B_{11}K_{1m} \\ \cdots & \cdots & \cdots \\ -A_{m1} - B_{mm}K_{m1} & \cdots & sI - A_{mm} - B_{mm}K_{mm} \end{bmatrix}P^{-1}. \end{aligned}$$

再由式(3.3.11)至式(3.3.13)易见可适当选取K_{ij} $(i,j = 1,2,\cdots,m)$, 使得

$$sI - A - BK = P\begin{bmatrix} s & -1 & 0 & \cdots & 0 \\ 0 & s & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & s & -1 \\ \beta_0 & \beta_1 & \cdots & \beta_{n-2} & s + \beta_{n-1} \end{bmatrix}P^{-1},$$

于是$\det(sI - A - BK) = \beta_0 + \beta_1 s + \cdots + \beta_{n-1}s^{n-1} + s^n$. □

推论 3.3.4 若$A \in \mathbb{R}^{n\times n}$和$B \in \mathbb{R}^{n\times m}$满足$(A,B)$能控, $s_1,\cdots,s_n \in \mathbb{C}$关于实轴对称(即若$s_i \notin \mathbb{R}$, 则存在$j \neq i$使得$s_j = \bar{s}_i$), 则存在$K \in \mathbb{R}^{m\times n}$使得$sI - A - BK$有特征根$s_1$, $\cdots$, s_n.

证明 设

$$\prod_{i=1}^{n}(s - s_i) = \beta_0 + \beta_1 s + \cdots + \beta_{n-1}s^{n-1} + s^n.$$

易见$\beta_0, \beta_1, \cdots, \beta_{n-1} \in \mathbb{R}$. 由推论3.3.3得证. □

3.4 满足秩条件的矩阵对的标准形

本节介绍几个满足一定秩条件的矩阵对的标准形.

定理 3.4.1 设$A \in \mathbb{F}^{m\times n}$, $B \in \mathbb{F}^{n\times p}$, 则

(i) $\mathrm{rank}(AB) = \mathrm{rank}A$的充要条件是存在可逆阵$P$, Q和R满足

$$A = P\mathrm{diag}(I_s, 0)Q^{-1}, \ B = Q\mathrm{diag}(I_t, 0)R,$$

其中$t = \mathrm{rank}B$, $s = \mathrm{rank}A$, $t \geqslant s$;

(ii) $\mathrm{rank}(AB) = \mathrm{rank}B$的充要条件是存在可逆阵$P$, Q和R满足

$$A = P\mathrm{diag}(I_s, 0)Q^{-1}, \ B = Q\mathrm{diag}(I_t, 0)R,$$

其中$t = \mathrm{rank}B$, $s = \mathrm{rank}A$, $t \leqslant s$.

证明 (i) 充分性显然, 下证必要性.

设B的等价分解为$B = Q_1\mathrm{diag}(I_t, 0)R_1$, 并设$A = \begin{bmatrix} A_1 & A_2 \end{bmatrix} Q_1^{-1}$, 其中$A_1 \in \mathbb{F}^{m\times t}$. 由$\mathrm{rank}(AB) = \mathrm{rank}A$得

$$\mathrm{rank}A_1 = \mathrm{rank}(AB) = \mathrm{rank}A,$$

于是存在$X \in \mathbb{F}^{t\times(n-t)}$使得$A_2 = A_1X$. 令

$$Q_2 = Q_1 \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix},$$

则

$$B = Q_2\mathrm{diag}(I_t, 0)R_1, \quad A = \begin{bmatrix} A_1 & 0 \end{bmatrix} Q_2^{-1}.$$

设A_1的等价分解为$A_1 = P\mathrm{diag}(I_s, 0)Q_3$, 则

$$A = \begin{bmatrix} P\mathrm{diag}(I_s, 0)Q_3 & 0 \end{bmatrix} Q_2^{-1} = P\mathrm{diag}(I_s, 0)\mathrm{diag}(Q_3, I)Q_2^{-1}$$

且

$$B = Q_2\mathrm{diag}(I_t, 0)R_1 = Q_2\mathrm{diag}(Q_3^{-1}, I)\mathrm{diag}(I_t, 0)\mathrm{diag}(Q_3, I)R_1.$$

令$Q = Q_2\mathrm{diag}(Q_3^{-1}, I)$及$R = \mathrm{diag}(Q_3, I)R_1$, 则$A = P\mathrm{diag}(I_s, 0)Q^{-1}$且$B = Q\mathrm{diag}(I_t, 0)R$.

(ii) 将(i)中的AB, A和B分别换成$(AB)^T, B^T$和A^T易得(ii). □

定理 3.4.2 设$A,B\in\mathbb{F}^{m\times n}$, 则$\mathrm{rank}(A+B)=\mathrm{rank}(A)+\mathrm{rank}(B)$的充要条件是存在可逆阵$P$和$Q$使得$A=P\mathrm{diag}(I_r,0)Q$且$B=P\mathrm{diag}(0,I_s)Q$, 其中$r+s\leqslant\min\{m,n\}$.

证明 充分性显然, 下证必要性.

由定理1.2.1和$\mathrm{rank}(A+B)=\mathrm{rank}(A)+\mathrm{rank}(B)$得

$$\mathcal{R}(A)\cap\mathcal{R}(B)=\{0\},\ \mathcal{N}(A)+\mathcal{N}(B)=\mathbb{F}^n. \tag{3.4.1}$$

设A的等价分解为$A=P_1\mathrm{diag}(I_r,0)Q_1$, 由$\mathcal{R}(A)\cap\mathcal{R}(B)=\{0\}$可设

$$B=P_1\begin{bmatrix}0&0\\C&D\end{bmatrix}Q_1,$$

其中$D\in\mathbb{F}^{(n-r)\times(n-r)}$, 于是

$$\mathcal{N}(A)=\left\{Q_1^{-1}\begin{bmatrix}0\\z\end{bmatrix}:z\in\mathbb{F}^{n-r}\right\},$$

$$\mathcal{N}(B)=\left\{Q_1^{-1}\begin{bmatrix}x\\y\end{bmatrix}:x\in\mathbb{F}^{r},\ y\in\mathbb{F}^{n-r},\ Cx+Dy=0\right\}.$$

使用$\mathcal{N}(A)+\mathcal{N}(B)=\mathbb{F}^n$推出: 对于任意的$x\in\mathbb{F}^r$, 存在$y\in\mathbb{F}^{n-r}$使得$Cx=-Dy=D(-y)$, 从而$\mathcal{R}(C)\subseteq\mathcal{R}(D)$, 即存在矩阵$Y$使得$C=DY$, 故

$$A=P_1\mathrm{diag}(I_r,0)\begin{bmatrix}I_r&0\\Y&I\end{bmatrix}Q_1,\ B=P_1\begin{bmatrix}0&0\\0&D\end{bmatrix}\begin{bmatrix}I_r&0\\Y&I\end{bmatrix}Q_1.$$

设D的等价分解为$D=P_2\mathrm{diag}(0,I_s)Q_2$, 并令

$$P=P_1\mathrm{diag}(I_r,P_2),\ Q=\mathrm{diag}(I_r,Q_2)\begin{bmatrix}I_r&0\\Y&I\end{bmatrix}Q_1,$$

则$A=P\mathrm{diag}(I_r,0)Q$且$B=P\mathrm{diag}(0,I_s)Q$. □

定理 3.4.3 设$A\in\mathbb{F}^{m\times n}$, $B\in\mathbb{F}^{n\times p}$, 则

$$\mathrm{rank}(AB)=\mathrm{rank}(A)+\mathrm{rank}(B)-n$$

的充要条件是存在可逆阵P, Q和S使得

$$A=P\mathrm{diag}(I_r,0)Q,\ B=Q^{-1}\mathrm{diag}(0,I_t)S,$$

其中$r+t\geqslant n$.

证明 充分性显然, 下证必要性.

由定理1.2.2和$\text{rank}(AB) = \text{rank}(A) + \text{rank}(B) - n$得

$$\mathcal{N}(A) \subseteq \mathcal{R}(B). \tag{3.4.2}$$

设B的等价分解为$B = T\text{diag}(0, I_t)S_1$, 并设$A = \begin{bmatrix} A_1 & A_2 \end{bmatrix} T^{-1}$, 则

$$\mathcal{R}(B) = \left\{ T \begin{bmatrix} 0 \\ z \end{bmatrix} : z \in \mathbb{F}^t \right\},$$

于是由式(3.4.2)得A_1列满秩, 否则, 存在非零的$x_0 \in \mathbb{F}^{n-t}$使得

$$T \begin{bmatrix} x_0 \\ 0 \end{bmatrix} \in \mathcal{N}(A),$$

与式(3.4.2)矛盾, 故存在矩阵C使得$P_1 := \begin{bmatrix} A_1 & C \end{bmatrix}$可逆, 从而

$$A = P_1 \begin{bmatrix} I_{n-t} & A_{12} \\ 0 & A_{22} \end{bmatrix} T^{-1}.$$

再使用式(3.4.2)推出$\mathcal{N}(A_{22}) \subseteq \mathcal{N}(A_{12})$, 于是

$$\mathcal{N}(A_{22}) = \mathcal{N}(A_{22}) \cap \mathcal{N}(A_{12}) = \mathcal{N}\left(\begin{bmatrix} A_{22} \\ A_{12} \end{bmatrix} \right),$$

即存在矩阵Z使得$A_{12} = ZA_{22}$. 令

$$P_2 = P_1 \begin{bmatrix} I & Z \\ 0 & I \end{bmatrix},$$

则$A = P_2\text{diag}(I_{n-t}, A_{22})T^{-1}$. 设$A_{22}$的等价分解为

$$A_{22} = P_3\text{diag}(I_{r+t-n}, 0)T_1,$$

$$P = P_2\text{diag}(I, P_3),\ Q = \text{diag}(I, T_1)T^{-1},\ S = \text{diag}(I, T_1)S_1,$$

则$A = P\text{diag}(I_r, 0)Q$且$B = Q^{-1}\text{diag}(0, I_t)S$. □

定理 3.4.4 设$E \in \mathbb{R}^{n\times n}$和$B \in \mathbb{R}^{n\times r}$满足

$$0 < \text{rank}E = n_0 < n, \quad \text{rank}B = r,$$

则存在正交阵T, V和可逆阵N使得

$$B = N \begin{bmatrix} 0_{n_1\times r} \\ I_r \\ 0 \end{bmatrix} T, E = N \begin{bmatrix} I_{n_1} & 0 \\ 0 & M \\ 0 & 0 \end{bmatrix} V, \tag{3.4.3}$$

其中$n_1 = \text{rank}\begin{bmatrix} E & B \end{bmatrix} - r$, $M = 0$或者M有形式:

$$M = \text{diag}(\Sigma, 0),\ \Sigma \in \mathbb{R}^{(n_0-n_1)\times(n_0-n_1)}\text{为正定对角阵.}$$

定理3.4.4的证明见附录C.

注记 3.4.1 定理3.4.4能够被用来解决广义系统的动态阶配置问题[10].

习 题

1. 已知$A=\begin{bmatrix} 2 & -2 & 1 \\ 1 & 1 & 0 \\ -1 & 3 & -1 \end{bmatrix}$, $B=\begin{bmatrix} 2 & 1 \\ 1 & 0 \\ -1 & -1 \end{bmatrix}$, 判定矩阵对$(A,B)$的能控性.

2. 设$E\in\mathbb{R}^{n\times n}$和$B\in\mathbb{R}^{n\times r}$满足

$$0<\text{rank}E=n_0<n,\quad \text{rank}B=r,$$

则下列说法等价:

(i) 存在$K\in\mathbb{R}^{r\times n}$使得$E+BK$可逆;

(ii) $\text{rank}\begin{bmatrix} E & B \end{bmatrix}=n$;

(iii) 存在正交阵T,V和可逆阵N使得

$$B=N\begin{bmatrix} 0_{(n-r)\times r} \\ I_r \end{bmatrix}T, E=N\begin{bmatrix} I_{n-r} & 0 \\ 0 & M \end{bmatrix}V,$$

其中$M=0$或者M有如下形式:

$$M=\text{diag}(\Sigma,0),\ (\Sigma\text{是实正定对角阵}).$$

(提示: 使用定理3.4.4.)

3. 设矩阵$A\in\mathbb{C}^{m\times n}$和$C\in\mathbb{C}^{p\times n}$满足$0<p=\text{rank}C\leqslant\text{rank}A=m<n$, 并设

$$s=\text{rank}\begin{bmatrix} C \\ A \end{bmatrix}-\text{rank}A,$$

则存在酉矩阵P和T, 可逆阵Q和实正定对角阵Σ使得

$$A=P\begin{bmatrix} I_m & 0 \end{bmatrix}Q,$$

$$C=T\left[\begin{array}{cc|cc} \Sigma & 0 & 0 & 0 \\ 0 & 0_{s\times(m-p+s)} & 0 & I_s \end{array}\right]Q.$$

(提示: 见参考文献[11].)

第4章　幂等矩阵与投影算子

投影在数学的许多分支中都有重要应用. 本章介绍从抽象线性空间到其子空间的投影算子, 由于投影算子在给定基下的阵是幂等矩阵, 故幂等矩阵和投影算子有很多相似的性质. 此外, 幂等矩阵在统计、优化等领域有重要应用.

4.1　幂等矩阵

定理 4.1.1　设$A \in \mathbb{F}^{n\times n}$, 则下列说法等价.

(i) $A^2 = A$(即A为幂等矩阵).

(ii) $(I_n - A)^2 = I_n - A$.

(iii) $\mathcal{R}(I_n - A) = \mathcal{N}(A)$.

(iv) $\operatorname{rank}A + \operatorname{rank}(I_n - A) = n$.

(v) 存在可逆阵$P \in \mathbb{F}^{n\times n}$使得$A = P\operatorname{diag}(I_r, 0)P^{-1}$, 其中$r = \operatorname{rank}A$.

证明　(i)⇒(ii). 显然.

(ii)⇒(iii). 由(ii)易见$(I_n - A)A = 0$, 于是$\mathcal{R}(A) \subseteq \mathcal{N}(I_n - A)$. 这和$\mathcal{N}(I_n - A) \subseteq \mathcal{R}(A)$一起推出(iii)成立.

(iii)⇒(iv). 由(iii)得

$$\operatorname{rank}(I_n - A) = \dim\mathcal{R}(I_n - A) = \dim\mathcal{N}(A) = n - \operatorname{rank}A,$$

即(iv)成立.

(iv)⇒(v). 设A的Fitting分解为$A = P\operatorname{diag}(D, N)P^{-1}$, 则

$$I - A = P\operatorname{diag}(I - D, I - N)P^{-1},$$

于是(iv)等价于

$$\text{rank}D+\text{rank}N+\text{rank}(I-D)+\text{rank}(I-N)=n.$$

使用D及$I-N$可逆推出$\text{rank}N+\text{rank}(I-D)=0$, 故$N=0$且$D=I$, 于是$A$的Fitting分解简化为$A=P\text{diag}(I,0)P^{-1}$.

(v)⇒(i) 显然. □

由定理4.1.1容易证明下面的推论(留给读者).

推论 4.1.1 设$A\in\mathbb{F}^{n\times n}$是幂等矩阵, $r=\text{rank}A$, 则

(i) A的特征值为1(r重)和0($n-r$重);

(ii) $\text{tr}A=r=\text{rank}A$;

(iii) $\mathcal{R}(A)\oplus\mathcal{N}(A)=\mathbb{F}^n$.

定理 4.1.2 设$k\geqslant 2$, $A_1,A_2,\cdots,A_k\in\mathbb{F}^{n\times n}$.

(a) $A_i^2=A_i$, $i=1,2,\cdots,k$;

(b) $A_iA_j=0$ $(i,j=1,2,\cdots,k,\ i\neq j)$, 并且$\text{rank}A_i=\text{rank}A_i^2$ $(i=1,2,\cdots,k)$;

(c) $A_1+A_2+\cdots+A_k$幂等;

(d) $\text{rank}(A_1+A_2+\cdots+A_k)=\text{rank}A_1+\text{rank}A_2+\cdots+\text{rank}A_k$.

则

(i) (b)⇒(d);

(ii) (c)和(d)同时成立的充要条件是(a)和(b)同时成立;

(iii) (a)至(c)中的任意两个可推出(a)至(d)中的其他两个.

证明 (i) 不妨设, $A_i\neq 0$, $i=1,2,\cdots,k$. 由

$$\text{rank}A_1=\text{rank}A_1^2$$

和Fitting分解知存在可逆阵P和D_1使得$A_1=P\text{diag}(D_1,0)P^{-1}$. 设

$$A_s=P\begin{bmatrix}A_{s1}&A_{s2}\\A_{s3}&A_{s4}\end{bmatrix}P^{-1},\ s=2,3,\cdots,k.$$

由$A_1A_s = A_sA_1 = 0$得

$$\begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} A_{s1} & A_{s2} \\ A_{s3} & A_{s4} \end{bmatrix} = \begin{bmatrix} A_{s1} & A_{s2} \\ A_{s3} & A_{s4} \end{bmatrix}\begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} = 0,$$

即

$$D_1A_{s1} = 0, \quad D_1A_{s2} = 0, \quad A_{s3}D_1 = 0.$$

使用D_1可逆推出$A_{s1} = 0, \quad A_{s2} = 0, \quad A_{s3} = 0$. 所以

$$A_s = P\text{diag}(0, B_s)P^{-1}, \quad s = 2, 3, \cdots, k,$$

其中$B_s = A_{s4}$. 由(b)知

$$\begin{cases} \text{rank}B_s = \text{rank}B_s^2, \ s = 2, \cdots, k, \\ B_sB_t = 0, \ s, t = 2, \cdots, k, s \neq t. \end{cases}$$

依此类推, 存在可逆阵Q使得$A_1 = Q\text{diag}(D_1, 0)Q^{-1}$且

$$A_i = Q\text{diag}(0_{p_i \times p_i}, D_i, 0, \cdots, 0)Q^{-1}, \ i = 2, \cdots, k, \tag{4.1.1}$$

其中$p_i = \sum\limits_{j=1}^{i-1} \text{rank}A_j$, D_i可逆. 故

$$A_1 + A_2 + \cdots + A_k = Q\text{diag}(D_1, D_2, \cdots, D_k, 0)Q^{-1}.$$

所以

$$\text{rank}(A_1 + A_2 + \cdots + A_k) = \text{rank}A_1 + \text{rank}A_2 + \cdots + \text{rank}A_k.$$

(ii) **充分性**. 由(i)的证明知(d)和式(4.1.1)成立. 再由(a)知$D_i^2 = D_i$, 于是$D_i = I$. 所以$A_1 + A_2 + \cdots + A_k = Q\text{diag}(I, 0)Q^{-1}$, 故$A_1 + A_2 + \cdots + A_k$幂等.

必要性. 设$A_0 = I_n - A_1 - A_2 - \cdots - A_k$, 由(c)知, A_0幂等. 又因

$$A_0 + A_1 + \cdots + A_k = I_n, \tag{4.1.2}$$

使用(d)和定理4.1.1推出

$$\begin{aligned} & \text{rank}A_i + \text{rank}(I - A_i) \\ = \ & \text{rank}A_i + \text{rank}\sum_{\substack{j=0 \\ j \neq i}}^{k} A_j \\ \leqslant \ & \text{rank}A_0 + \text{rank}A_1 + \cdots + \text{rank}A_k \\ = \ & \text{rank}A_0 + \text{rank}(A_1 + A_2 + \cdots + A_k) \\ = \ & \text{rank}A_0 + \text{rank}(I_n - A_0) \\ = \ & n. \end{aligned}$$

另一方面, $n=\mathrm{rank}\,(A_i+(I_n-A_i))\leqslant \mathrm{rank}A_i+\mathrm{rank}(I_n-A_i)$. 故

$$\mathrm{rank}A_0+\mathrm{rank}A_1+\cdots+\mathrm{rank}A_k=\mathrm{rank}A_i+\mathrm{rank}(I_n-A_i)=n, \quad (4.1.3)$$

再由定理4.1.1推出$A_i^2=A_i$, 亦即(a)成立. 故

$$\mathrm{rank}A_i=\mathrm{rank}A_i^2,\ i=1,2,\cdots,k.$$

由式(4.1.2)和式(4.1.3)得

$$\mathcal{R}(A_0)\oplus\mathcal{R}(A_1)\oplus\cdots\oplus\mathcal{R}(A_k)=\mathbb{F}^n. \quad (4.1.4)$$

另一方面, 由式(4.1.2)及(a)得

$$A_0A_j+\cdots+A_{j-1}A_j+A_{j+1}A_j+\cdots+A_kA_j=0,\ j=1,2,\cdots,k,$$

从而

$$A_0A_jx+A_1A_jx+\cdots+A_{j-1}A_jx+A_{j+1}A_jx+\cdots+A_kA_jx=0,$$
$$\forall\ x\in\mathbb{F}^n,\ j=1,2,\cdots,k.$$

使用式(4.1.4)和零元的表出唯一性得

$$A_iA_j=0,\ \forall\ i\neq j.$$

故(b)成立.

(iii) 由(i)和(ii)知, 只需证: 若(a)和(c)同时成立, 则(d)成立. 事实上, 设$A=A_1+A_2+\cdots+A_k$, 则$\mathrm{tr}A=\mathrm{tr}A_1+\mathrm{tr}A_2+\cdots+\mathrm{tr}A_k$. 由(a), (c)及推论4.1.1得$\mathrm{rank}A=\mathrm{rank}A_1+\mathrm{rank}A_2+\cdots+\mathrm{rank}A_k$, 即(d)成立. □

定理 4.1.3 设$A,B\in\mathbb{F}^{n\times n}$幂等. 证明

(i) $AB=BA=0$的充要条件是$A+B$幂等;

(ii) 若$AB=BA$, 则存在可逆阵$P\in\mathbb{F}^{n\times n}$使得

$$P^{-1}AP=\mathrm{diag}(I_r,0),\ P^{-1}BP=\mathrm{diag}\left(\begin{bmatrix} I_s & 0\\ 0 & 0\end{bmatrix},\begin{bmatrix} 0 & 0\\ 0 & I_t\end{bmatrix}\right);$$

(iii) 若$A+B$幂等, 则存在可逆阵$P\in\mathbb{F}^{n\times n}$使得

$$P^{-1}AP=\mathrm{diag}(I_r,0),\ P^{-1}BP=\mathrm{diag}(0,I_t),$$

其中$r+t\leqslant n$.

证明 (i) 由A, B幂等得$(A+B)^2 = A^2 + AB + BA + B^2 = A + AB + BA + B$, 于是只需证$AB = BA = 0$的充要条件是

$$AB + BA = 0. \tag{4.1.5}$$

若式(4.1.5)成立, 在式(4.1.5)两边分别左、右乘A得

$$AB + ABA = 0 = ABA + BA,$$

从而$AB = BA$. 再使用式(4.1.5)推出

$$AB = BA = 0.$$

反之, 若$AB = BA = 0$, 则易见式(4.1.5)成立. 故式(4.1.5)成立的充要条件是$AB = BA = 0$.

总之, $AB = BA = 0$的充要条件是$A + B$幂等.

(ii) 由A幂等知存在可逆阵$P_1 \in \mathbb{F}^{n\times n}$使得$P_1^{-1}AP_1 = \mathrm{diag}(I_r, 0)$. 使用$AB = BA$得到

$$P_1^{-1}BP_1 = \mathrm{diag}(B_1, B_2).$$

再由B幂等推出B_1和B_2均幂等, 于是存在可逆阵P_2和P_3使得

$$P_2^{-1}B_1P_2 = \mathrm{diag}(I_s, 0),\ P_3^{-1}B_2P_3 = \mathrm{diag}(0, I_t).$$

令$P = P_1\mathrm{diag}(P_2, P_3)$得证.

(iii) 由A, B, $A + B$幂等和(i)得$AB = BA = 0$. 再由(ii)得证. □

4.2 投影算子与投影矩阵

设V是$\mathbb{F}$上的n维线性空间, L和M是V的子空间, 且$L \oplus M = V$, 则V中的每个向量x可以唯一地表示成

$$x = x_L + x_M,\ x_L \in L,\ x_M \in M.$$

称x_L为x沿M在L上的投影, x_M为x沿L在M上的投影. 称将V中的任一向量x映为x_L的变换为V的沿M到L的投影算子, 记作$P^{L,M}$. 由此,

$$x_L = P^{L,M}(x),\ x_M = P^{M,L}(x),$$

于是

$$x = P^{L,M}(x) + P^{M,L}(x) = (P^{L,M} + P^{M,L})(x),\ \forall x \in V,$$

因而$P^{L,M} + P^{M,L} = 1_V$, 其中1_V表示V上的恒等映射. 进一步地, 投影算子能被刻画如下.

定理 4.2.1 设V是$\mathbb{F}$上的n维线性空间, σ是V上的变换, 则σ是投影算子的充要条件是σ是线性的且$\sigma^2=\sigma$.

证明 **必要性**. 设V的子空间V_1和V_2满足$V=V_1\oplus V_2$, 并且$\sigma=P^{V_1,V_2}$. 任取$a,b\in\mathbb{F}$和$x,y\in V$, 则x,y可唯一表示为

$$x=x_1+x_2,\ y=y_1+y_2,$$

其中$x_1,y_1\in V_1$, $x_2,y_2\in V_2$. 由投影算子的定义得

$$\begin{aligned}\sigma(ax+by)=&\ \sigma((ax_1+by_1)+(ax_2+by_2))\\=&\ ax_1+by_1\\=&\ a\sigma(x)+b\sigma(y)\end{aligned}$$

及

$$\sigma^2(x)=\sigma(\sigma(x))=\sigma(x_1)=x_1=\sigma(x).$$

故σ是线性的且$\sigma^2=\sigma$.

充分性. 既然σ是线性的且$\sigma^2=\sigma$, 由推论4.1.1知$\mathrm{Im}\sigma\oplus\ker\sigma=V$. 任取$x\in V$, 则$x$有唯一分解

$$x=y+z,\ y\in\mathrm{Im}\,\sigma,z\in\ker\sigma.$$

令$w\in V$满足$y=\sigma(w)$, 则

$$\sigma(x)=\sigma(y+z)=\sigma(y)+\sigma(z)=\sigma(y)=\sigma(\sigma(w))=\sigma^2(w)=\sigma(w)=y,$$

故σ是沿$\ker\sigma$到$\mathrm{Im}\,\sigma$的投影算子. □

现在讨论$P^{L,M}$在V的基下的矩阵表示. 设$z_1,\cdots,z_r$是L的基, $z_{r+1},\cdots,z_n$是M的基, 则$z_1,\cdots,z_r,z_{r+1},\cdots,z_n$是$V$的基. 易见

$$P^{L,M}(z_1,\cdots,z_n)=(z_1,\cdots,z_n)\mathrm{diag}(I_r,0),$$

故$P^{L,M}$在V的基$z_1,\cdots,z_n$下的矩阵表示为$\mathrm{diag}(I_r,0)$. 若$w_1,\cdots,w_n$是V的另一组基, 则存在可逆阵T使得$(z_1,\cdots,z_n)=(w_1,\cdots,w_n)T$, 于是

$$P^{L,M}(w_1,\cdots,w_n)T=(w_1,\cdots,w_n)T\mathrm{diag}(I_r,0),$$

因而$P^{L,M}$在V的基$\omega_1,\cdots,\omega_n$下的矩阵表示为$T\mathrm{diag}(I_r,0)T^{-1}$.

当取定V的基后, 用$P_{L,M}$表示投影算子$P^{L,M}$在该基下的矩阵(称为投影矩阵), 易见$P_{L,M}$是幂等矩阵, 并且$P_{L,M}+P_{M,L}=I_n$. 注意到投影矩阵和投影算子的对应关系, 组合定理4.1.1和定理4.2.1得到下面的结论.

定理 4.2.2 设V是$\mathbb{F}$上的n维线性空间, σ是V上的线性变换, 则下列说法等价:

(i) σ是投影算子.

(ii) $\sigma^2=\sigma$.

(iii) $(1_V-\sigma)^2=1_V-\sigma$.

(iv) $\mathrm{Im}\,(1_V-\sigma)=\ker\sigma$.

下面我们介绍投影矩阵的一些性质.

定理 4.2.3 设$A\in\mathbb{F}^{n\times n}$幂等, 则$A$为投影矩阵, 并且$A=P_{\mathcal{R}(A),\mathcal{N}(A)}$.

证明 定义线性变换$\sigma:\mathbb{F}^n\to\mathbb{F}^n$为

$$\sigma(x)=Ax,\ \forall x\in\mathbb{F}^n.$$

由A幂等知$\sigma^2=\sigma$, 于是由定理4.2.1的证明得$\sigma=P^{\mathrm{Im}\sigma,\ker\sigma}$是投影算子. 再使用$\sigma$在$\mathbb{F}^n$的自然基下的阵为$A$推出$A$为投影矩阵, 并且$A=P_{\mathcal{R}(A),\mathcal{N}(A)}$. □

定理 4.2.4 设$P_1,P_2\in\mathbb{F}^{n\times n}$是两个投影矩阵, 则

(i) P_1+P_2为投影矩阵的充要条件是$P_1P_2=P_2P_1=0$;

(ii) 当$P_1P_2=P_2P_1=0$时有$P_1+P_2=P_{L,M}$, 其中$L=\mathcal{R}(P_1)\oplus\mathcal{R}(P_2)$, $M=\mathcal{N}(P_1)\cap\mathcal{N}(P_2)$.

证明 既然投影矩阵是幂等矩阵, 组合定理4.2.3和定理4.1.3得到. □

定理 4.2.5 设$P_1,P_2\in\mathbb{F}^{n\times n}$是两个投影矩阵, 则

(i) P_1-P_2为投影矩阵的充要条件是$P_1P_2=P_2P_1=P_2$;

(ii) 当$P_1P_2=P_2P_1=P_2$时有$P_1-P_2=P_{L,M}$, 其中$L=\mathcal{R}(P_1)\cap\mathcal{N}(P_2)$, $M=\mathcal{N}(P_1)\oplus\mathcal{R}(P_2)$.

证明 (i) 既然$I-P_1$是投影矩阵, 由定理4.2.4(i)得到

$$\begin{aligned}P_1-P_2\text{为投影矩阵}&\Leftrightarrow I-P_1+P_2\text{为投影矩阵}\\&\Leftrightarrow (I-P_1)P_2=P_2(I-P_1)=0\\&\Leftrightarrow P_1P_2=P_2P_1=P_2.\end{aligned}$$

(ii) 由$P_1P_2=P_2P_1=P_2$得$(I-P_1)P_2=P_2(I-P_1)=0$. 使用定理4.1.1和定理4.2.4(ii)推出

$$I-P_1+P_2=P_{\mathcal{R}(I-P_1)\oplus\mathcal{R}(P_2),\mathcal{N}(I-P_1)\cap\mathcal{N}(P_2)}=P_{\mathcal{N}(P_1)\oplus\mathcal{R}(P_2),\mathcal{R}(P_1)\cap\mathcal{N}(P_2)},$$

于是$P_1-P_2=P_{\mathcal{R}(P_1)\cap\mathcal{N}(P_2),\mathcal{N}(P_1)\oplus\mathcal{R}(P_2)}$. □

定理 4.2.6 设$A \in \mathbb{F}^{n\times n}$, $\mathbb{F}^n$的子空间L和M满足$\mathbb{F}^n = L \oplus M$, 则

$$P_{L,M}A = A$$

的充要条件是

$$\mathcal{R}(A) \subset L.$$

证明

$$\begin{aligned} P_{L,M}A = A \Leftrightarrow &\ P_{L,M}a_j = a_j,\ j = 1,2,\cdots,n \\ \Leftrightarrow &\ a_j \in L,\ j = 1,2,\cdots,n \\ \Leftrightarrow &\ \mathcal{R}(A) \subset L, \end{aligned}$$

其中a_j是A的第j列. □

定理 4.2.7 设$\mathbb{C}^n$的子空间L和M满足$\mathbb{C}^n = L \oplus M$, 则

(i) $P^*_{L,M} = P_{M^\perp,L^\perp}$;

(ii) 若$A \in \mathbb{C}^{n\times n}$, 则$AP_{L,M} = A$的充要条件是$\mathcal{N}(A) \supset M$.

证明 (i) 由$\mathcal{R}(P^*_{L,M}) = \mathcal{N}(P_{L,M})^\perp = M^\perp$和$\mathcal{N}(P^*_{L,M}) = \mathcal{R}(P_{L,M})^\perp = L^\perp$得$P^*_{L,M} = P_{M^\perp,L^\perp}$.

(ii) 由定理4.2.6知

$$P_{M^\perp,L^\perp}A^* = A^* \Leftrightarrow \mathcal{R}(A^*) \subset M^\perp \Leftrightarrow \mathcal{N}(A)^\perp \subset M^\perp \Leftrightarrow M \subset \mathcal{N}(A).$$

使用(i)推出$AP_{L,M} = A$的充要条件是$\mathcal{N}(A) \supset M$. □

4.3 正交投影矩阵

设L是$\mathbb{C}^n$的子空间, 称V的沿着$L^\perp$到L的投影算子$P^{L,L^\perp}$为正交投影算子, 简记为P^L. 称正交投影算子在某组标准正交基下对应的矩阵$P_{L,L^\perp}$为正交投影矩阵, 简记为P_L.

定理 4.3.1 下列说法等价:

(i) P是正交投影矩阵.

(ii) P是Hermite幂等矩阵.

(iii) 存在酉矩阵U使得$P = U\mathrm{diag}(I_r, 0)U^*$, 其中$r = \mathrm{rank}P$.

(iv) $P = U_1U_1^*$, 其中$U_1 \in \mathbb{C}^{n\times r}$满足$U_1^*U_1 = I_r$.

证明 留给读者. □

定理 4.3.2 若A是Hermite幂等矩阵, 则$P_{\mathcal{R}(A)}A = A$且$P_{\mathcal{R}(A)}x = 0$, $\forall x \in \mathcal{R}(A)^{\perp}$.

证明 组合定理4.2.6($L = \mathcal{R}(A)$)和推论4.1.1得到. □

定理 4.3.3 设P_1和P_2是正交投影矩阵, 则

(i) P_1P_2为正交投影矩阵的充要条件是$P_1P_2 = P_2P_1$;

(ii) 当$P_1P_2 = P_2P_1$时有$P_1P_2 = P_{L,M}$, 其中$L = \mathcal{R}(P_1) \cap \mathcal{R}(P_2)$, $M = \mathcal{N}(P_1) + \mathcal{N}(P_2)$.

证明 (i) **充分性**. 由P_1和P_2是正交投影矩阵得$P_i^2 = P_i$且$P_i^* = P_i$, $i = 1, 2$, 于是

$$(P_1P_2)^* = P_2^*P_1^* = P_2P_1 = P_1P_2,$$
$$(P_1P_2)^2 = P_1P_2P_1P_2 = P_1(P_1P_2)P_2 = P_1^2P_2^2 = P_1P_2.$$

故P_1P_2为正交投影矩阵.

必要性. 由P_1P_2是正交投影矩阵得$(P_1P_2)^* = P_1P_2$, 于是

$$P_1P_2 = (P_1P_2)^* = P_2^*P_1^*.$$

使用P_1和P_2是正交投影矩阵推出$P_1P_2 = P_2P_1$.

(ii) 由定理4.1.3(ii)易见. □

习 题

1. 设$A \in \mathbb{F}^{n\times n}$, 则下列说法等价:

   (i) $A^2 = A$;

   (ii) $\mathcal{R}(I_n - A) \oplus \mathcal{R}(A) = \mathbb{F}^n$;

   (iii) $\mathcal{R}(I_n - A) \cap \mathcal{R}(A) = \{0\}$.

2. 举例说明定理4.1.2中的(a)和(d)不能推出(b)或(c), 并且(b)和(d)也不能推出(a)或(c).

3. 设$A \in \mathbb{F}^{n\times n}$, 则$A^2 = A$的充要条件是$Ax = x$, $\forall x \in \mathcal{R}(A)$.

4. 若A的满秩分解为$A = FG$, 则$A^2 = A$的充要条件是$GF = I$.

5. 设V是$\mathbb{F}$上的n维线性空间, σ是V上的线性变换, 则下列说法等价:

   (i) σ是投影算子;

   (ii) $\mathrm{Im}\,(1_V - \sigma) \oplus \mathrm{Im}\,(\sigma) = V$;

   (iii) $\mathrm{rank}\sigma + \mathrm{rank}(1_V - \sigma) = n$.

6. 若$\mathbb{C}^n = L \oplus M$, 则存在唯一的幂等矩阵$P \in \mathbb{C}^{n\times n}$使得$\mathcal{R}(P) = L, \mathcal{N}(P) = M$.

7. 设P是正交投影矩阵, 证明: $P \geqslant 0$.

第5章　向量范数

本章介绍实数域或复数域上的线性空间上的范数的定义、例子和性质. 主要给出范数的等价性、相容性和从属范数、谱半径、条件数、矩阵测度等概念和基本结论. 既然范数是长度的推广, 它在许多领域中都有重要应用.

5.1　向量范数的定义和例子

定义 5.1.1　设V是$\mathbb{F}$上的线性空间. 若V上的实值函数$\|\cdot\|$满足下面的(i)~(iii), 称$\|\cdot\|$是V上的范数.

(i) (非负性) $\|x\| \geqslant 0,\ \forall\, x \in V$, 并且$\|x\| = 0 \Leftrightarrow x = 0$;

(ii) (齐次性) $\|kx\| = |k|\|x\|, \forall\, k \in \mathbb{F}, x \in V$;

(iii) (三角不等式) $\|x + y\| \leqslant \|x\| + \|y\|, \forall\, x, y \in V$.

当$\mathbb{F} = \mathbb{R}$且$V = \mathbb{C}$时, 若定义$\|x\|$为复数x的模长, 易见$\|\cdot\|$满足上面定义的(i)~(iii), 因而可将向量范数看成是长度的推广.

引理 5.1.1 (Hölder 不等式)　[12] 若$a_i, b_i \in \mathbb{F}$, $i = 1, 2, \cdots, n$, $p, q \in \mathbb{R}$满足$p, q > 1$且$\frac{1}{p} + \frac{1}{q} = 1$, 则

$$\sum_{i=1}^{n} |a_i b_i| \leqslant \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |b_i|^q\right)^{1/q}.$$

例 5.1.1　设$p \in \mathbb{R}$满足$p \geqslant 1$, 定义$\mathbb{F}^n$上的实值函数$\|\cdot\|_p$如下:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p},\ \forall\, x = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^T \in \mathbb{F}^n. \tag{5.1.1}$$

现在证明$\|\cdot\|_p$是$\mathbb{F}^n$上的范数(称为向量l_p-范数). 事实上, 显然$\|x\|_p$满足定义5.1.1中的(i)和(ii), 因而只需证定义5.1.1中的(iii)成立. 事实上, 当$p=1$时, 式(5.1.1)简化成

$$\|x\|_1=\sum_{i=1}^n|x_i|,\ \forall x=\left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right]^T\in\mathbb{F}^n,$$

于是(iii)成立. 当$p>1$时, 设$q=\frac{p}{p-1}$, 由式(5.1.1)和引理5.1.1有

$$\begin{aligned}
\|x+y\|_p^p= & \sum_{i=1}^n|x_i+y_i|^p \\
= & \sum_{i=1}^n\left(|x_i+y_i||x_i+y_i|^{p-1}\right) \\
\leqslant & \sum_{i=1}^n\left(|x_i||x_i+y_i|^{p-1}\right)+\sum_{i=1}^n\left(|y_i||x_i+y_i|^{p-1}\right) \\
\leqslant & \left(\sum_{i=1}^n|x_i|^p\right)^{1/p}\cdot\left(\sum_{i=1}^n|x_i+y_i|^{(p-1)q}\right)^{1/q} \\
& +\left(\sum_{i=1}^n|y_i|^p\right)^{1/p}\cdot\left(\sum_{i=1}^n|x_i+y_i|^{(p-1)q}\right)^{1/q} \\
= & \left(\sum_{i=1}^n|x_i+y_i|^{(p-1)q}\right)^{1/q}(\|x\|_p+\|y\|_p) \\
= & \left(\sum_{i=1}^n|x_i+y_i|^p\right)^{1-\frac{1}{p}}(\|x\|_p+\|y\|_p) \\
= & \frac{\sum_{i=1}^n|x_i+y_i|^p}{\left(\sum_{i=1}^n|x_i+y_i|^p\right)^{1/p}}\cdot(\|x\|_p+\|y\|_p) \\
= & \frac{\sum_{i=1}^n|x_i+y_i|^p}{\|x+y\|_p}\cdot(\|x\|_p+\|y\|_p),\ \forall x,y\in\mathbb{F}^n,
\end{aligned}$$

故定义5.1.1的(iii)成立.

通常简称向量l_1-范数为向量1−范数. 当$p=2$时, 式(5.1.1)简化成

$$\|x\|_2=\left(\sum_{i=1}^n|x_i|^2\right)^{\frac{1}{2}},\ \forall x=\left[\begin{array}{ccc} x_1 & \cdots & x_n \end{array}\right]^T\in\mathbb{F}^n.$$

通常简称向量l_2-范数为向量2−范数.

例 5.1.2 考虑由式(5.1.1)定义的$\mathbb{F}^n$上的实值函数$\|\cdot\|_p$, 易见

$$\begin{aligned}\lim_{p\to+\infty}\|x\|_p = & \lim_{p\to+\infty}\left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p} \\ = & \max_{1\leqslant i\leqslant n}|x_i|\lim_{p\to+\infty}\left(\sum_{i=1}^{n}|y_i|^p\right)^{1/p}, \ \forall x\in\mathbb{F}^n, x\neq 0,\end{aligned}$$

其中$y_i = \dfrac{x_i}{\max\limits_{1\leqslant i\leqslant n}|x_i|}$. 使用

$$1\leqslant\sum_{i=1}^{n}y_i\leqslant\left(\sum_{i=1}^{n}1^p\right)^{1/p}\left(\sum_{i=1}^{n}y_i^p\right)^{1/p}\leqslant n^{2/p}$$

推出

$$\lim_{p\to+\infty}\|x\|_p = \max_{1\leqslant i\leqslant n}|x_i|, \ \forall x\in\mathbb{F}^n.$$

进而, 容易验证$\lim\limits_{p\to+\infty}\|x\|_p$是$\mathbb{F}^n$上的范数, 称为向量$\infty-$范数, 记作$\|\cdot\|_\infty$.

注记 5.1.1 例5.1.1和例5.1.2表明: 对于任意的实数$p\geqslant 1$(p可取做无穷大), 由式(5.1.1)定义的$\mathbb{F}^n$上的实值函数$\|\cdot\|_p$均是向量范数. 然而, 当$0<p<1$时, 由式(5.1.1)定义的$\|\cdot\|_p$不是向量范数. 事实上, 取$x=e_1$及$y=e_2$, 则$\|x+y\|_p = 2^{\frac{1}{p}}$, $\|x\|_p = \|y\|_p = 1$, 于是$\|x+y\|_p > \|x\|_p + \|y\|_p$, 故当$0<p<1$时, 由5.1.1定义的$\|\cdot\|$不是向量范数.

例 5.1.3 设A是复Hermite正定矩阵, 定义$\mathbb{C}^n$上的实值函数

$$\|x\| = (x^*Ax)^{\frac{1}{2}}, \ \forall x\in\mathbb{C}^n.$$

容易验证$\|\cdot\|$是$\mathbb{C}^n$上的范数. (留给读者)

定理 5.1.1 设$\|\cdot\|$是$\mathbb{F}^n$上的范数, $T\in\mathbb{F}^{n\times n}$, 定义$\mathbb{F}^n$上的实值函数

$$\|x\|_T = \|Tx\|, \ \forall x\in\mathbb{F}^n,$$

则$\|\cdot\|_T$是$\mathbb{F}^n$上的范数的充要条件是T可逆.

证明 **必要性**. 由$\|\cdot\|_T$和$\|\cdot\|$是$\mathbb{F}^n$的范数得

$$x=0 \ \Leftrightarrow \ \|x\|_T=0 \ \Leftrightarrow \ \|Tx\|=0 \ \Leftrightarrow \ Tx=0 \Leftrightarrow T\text{可逆}.$$

充分性. 显然$\|\cdot\|_T$满足定义5.1.1中的(i)和(ii). 对任意的$x, y \in \mathbb{F}^n$有

$$\|x+y\|_T = \|T(x+y)\| = \|Tx+Ty\| \leqslant \|Tx\| + \|Ty\| = \|x\|_T + \|y\|_T,$$

因而$\|\cdot\|_T$是$\mathbb{F}^n$上的范数. □

既然$\mathbb{F}^{n\times n}$中的可逆阵有无穷多个, 故定理5.1.1表明: 一旦定义了$\mathbb{F}^n$上的一个范数, 则可以定义$\mathbb{F}^n$上无穷多个范数.

例 5.1.4 定义$\mathbb{C}^{n\times n}$上的实值函数

$$\|A\| = (\mathrm{tr}(A^*A))^{\frac{1}{2}},\ \forall A \in \mathbb{C}^{n\times n},$$

容易验证$\|A\|$是$\mathbb{C}^{n\times n}$上的范数.

例 5.1.5 设V是$\mathbb{F}$上的线性空间, $\|\cdot\|_a$和$\|\cdot\|_b$是V上的范数, 定义V上的实值函数$\|\cdot\|$为

$$\|x\| = \max\{\|x\|_a, \|x\|_b\},\ \forall x \in V,$$

则$\|x\|$是V上的范数.(请读者自己验证)

5.2 范数的等价性

例5.1.3表明可通过正定阵构造向量范数, 而定理5.1.1和例5.1.5表明可通过已知向量范数来构造新的向量范数. 这些例子表明同一个线性空间上可以定义无穷多个范数. 然而, 下面将证明有限维线性空间上的任意两个范数都是等价的. 这使得在研究许多问题(例如: 矩阵序列的收敛性问题)时可以任意选择一种范数使用, 也可以根据实际需要来定义新的范数.

定理 5.2.1 设V是$\mathbb{F}$上的线性空间, $x^{(1)}, x^{(2)}, \cdots, x^{(n)} \in V$, $\|\cdot\|$ 是V上的范数, 定义$\mathbb{F}^n$上的实值函数

$$g(z) = \|z_1x^{(1)} + z_2x^{(2)} + \cdots + z_nx^{(n)}\|,\ \forall z = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix}^T \in \mathbb{F}^n,$$

则g关于z一致连续.

证明 对于任意的$\varepsilon > 0$, 存在$\delta = \dfrac{\varepsilon}{\sum\limits_{i=1}^{n}\|x^{(i)}\|}$ 使得当$u, v \in \mathbb{F}^n$满足$\|u -$

$v\|_\infty < \delta$时, 有

$$\begin{aligned} |g(u)-g(v)| = & \ \left|\|u_1x^{(1)}+\cdots+u_nx^{(n)}\| - \|v_1x^{(1)}+\cdots+v_nx^{(n)}\|\right| \\ \leqslant & \ \left\|\sum_{i=1}^{n}(u_i-v_i)\,x^{(i)}\right\| \\ \leqslant & \ \sum_{i=1}^{n}|u_i-v_i|\|x^{(i)}\| \\ \leqslant & \ \|u-v\|_\infty \sum_{i=1}^{n}\|x^{(i)}\| \\ \leqslant & \ \varepsilon, \end{aligned}$$

故g关于z一致连续. □

引理 5.2.1 (Weierstrass**定理**) [13] 设S是有限维向量空间V中的紧集, f是S上的实值连续函数, 则存在$x_{\min}\in S$和$x_{\max}\in S$满足

$$f(x_{\min})\leqslant f(x)\leqslant f(x_{\max}),\ \forall x\in S.$$

定理 5.2.2 设V是$\mathbb{F}$上的有限维线性空间, $x^{(1)}$, $x^{(2)}$, $\cdots$, $x^{(n)}$是V的基, f_1和f_2是V上的两个实值函数, 若下面的(a)~(c)成立, 则存在正常数c_m和c_M使得

$$c_mf_1(x)\leqslant f_2(x)\leqslant c_Mf_1(x),\ \forall x\in V.$$

(a) $f_i(x)\geqslant 0$, $\forall x\in V$, 并且$f_i(x)=0\Leftrightarrow x=0$, $i=1,2$;

(b) $f_i(kx)=|k|f_i(x)$, $\forall k\in\mathbb{F},\ x\in V,\ \ i=1,2$;

(c) $f_i(x(z))$关于z连续, $i=1,2$, 其中

$$x(z)=z_1x^{(1)}+z_2x^{(2)}+\cdots+z_nx^{(n)},\ \forall z=\begin{bmatrix} z_1 & \cdots & z_n\end{bmatrix}^T\in\mathbb{F}^n.$$

证明 令$S=\{z\in\mathbb{F}^n \mid \|z\|_2=1\}$, 则$S$是$\mathbb{F}^n$中的紧集. 令

$$g\left(z\right)=\frac{f_2\left(x\left(z\right)\right)}{f_1\left(x\left(z\right)\right)},\ \ \forall\ z\in S.$$

由(a)和(c)知, g是S上的实值连续函数. 由引理5.2.1知存在$z_{\min}, z_{\max}\in S$使得

$$c_m:=g\left(z_{\min}\right)\leqslant g\left(z\right)\leqslant g\left(z_{\max}\right):=c_M,\ \forall\ z\in S,$$

于是

$$c_m f_1(x(z)) \leqslant f_2(x(z)) \leqslant c_M f_1(x(z)),\ \forall\ z \in S.$$

任取$0 \neq y \in \mathbb{F}^n$, 则$\frac{y}{\|y\|_2} \in S$. 故

$$c_m f_1\left(x\left(\frac{y}{\|y\|_2}\right)\right) \leqslant f_2\left(x\left(\frac{y}{\|y\|_2}\right)\right) \leqslant c_M f_1\left(x\left(\frac{y}{\|y\|_2}\right)\right),\ \forall 0 \neq y \in \mathbb{F}^n.$$

由(b)得

$$\frac{c_m}{\|y\|_2} f_1(x(y)) \leqslant \frac{1}{\|y\|_2} f_2(x(y)) \leqslant \frac{c_M}{\|y\|_2} f_1(x(y)),\ \forall 0 \neq y \in \mathbb{F}^n.$$

注意到y的任意性及$x^{(1)}, x^{(2)}, \cdots, x^{(n)}$是$V$的基, 结论得证. □

在定理5.2.2中将f_1和f_2分别取做V上的范数$\|\cdot\|_a$和$\|\cdot\|_b$可得范数的等价性定理如下:

定理 5.2.3 (范数等价性定理) 设V是$\mathbb{F}$上的有限维线性空间, $\|\cdot\|_a$和$\|\cdot\|_b$是V上的两个范数, 则存在正常数c_m和c_M使得

$$c_m\|x\|_a \leqslant \|x\|_b \leqslant c_M\|x\|_a,\ \forall x \in V.$$

5.3 矩阵范数

本节介绍一类特殊的向量范数, 即将定义5.1.1中的线性空间V取做矩阵空间$\mathbb{F}^{m\times n}$, 称这类特殊的向量范数为矩阵范数. 为了清楚地刻画矩阵范数与$\mathbb{F}^n$上的范数之间的本质联系, 首先定义拉直映射$\mathrm{Vec}: \mathbb{F}^{m\times n} \longrightarrow \mathbb{F}^{mn}$如下

$$\mathrm{Vec}: \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \mapsto \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix},\ \forall \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \in \mathbb{F}^{m\times n}.$$

易见Vec是从矩阵空间$\mathbb{C}^{m\times n}$到$\mathbb{C}^{mn}$的同构映射. 因而可模仿$\mathbb{C}^n$上的1−范数、∞−范数、2−范数来定义三种矩阵范数如下:

(i) (m_1 − 范数) $\|A\|_{m_1} = \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{ij}|,\ \forall\ A = [a_{ij}] \in \mathbb{F}^{m\times n}$;

(ii) (m_∞ − 范数) $\|A\|_{m_\infty} = n \max\limits_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}}|a_{ij}|,\ \forall\ A = [a_{ij}] \in \mathbb{F}^{m\times n}$;

(iii) (m_2 − 范数) $\|A\|_{m_2} = \left(\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}},\ \forall\ A = [a_{ij}] \in \mathbb{F}^{m\times n}$.

5.3.1 范数的相容性

定义 5.3.1 设$\|\cdot\|_a, \|\cdot\|_b$和$\|\cdot\|_c$分别是$\mathbb{F}^{m\times n}, \mathbb{F}^{m\times p}$和$\mathbb{F}^{p\times n}$上的范数,若

$$\|AB\|_a \leqslant \|A\|_b\|B\|_c,\ \forall\ A \in \mathbb{F}^{m\times p}, B \in \mathbb{F}^{p\times n},$$

称$\|\cdot\|_a$相容于$\|\cdot\|_b$和$\|\cdot\|_c$. 特别地, 当$\|\cdot\|_a = \|\cdot\|_c$时, 称$\|\cdot\|_a$和$\|\cdot\|_b$相容.

定义 5.3.2 若$\mathbb{F}^{n\times n}$上的范数$\|\cdot\|$满足

$$\|AB\| \leqslant \|A\|\|B\|,\ \forall\ A, B \in \mathbb{F}^{n\times n},$$

称$\|\cdot\|$是$\mathbb{F}^{n\times n}$上的相容范数.

下面给出几个关于范数相容性的定理.

定理 5.3.1 范数$\|\cdot\|_1$相容于范数$\|\cdot\|_{m_1}$和$\|\cdot\|_\infty$.

证明 任取$A = [a_{ij}] \in \mathbb{F}^{m\times n}$, $x = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^T \in \mathbb{F}^n$, 易见

$$\begin{aligned} \|Ax\|_1 = & \sum_{i=1}^m \left| \sum_{k=1}^n a_{ik}x_k \right| \\ \leqslant & \sum_{i=1}^m \sum_{k=1}^n |a_{ik}x_k| \\ \leqslant & \max_{1\leqslant k\leqslant n} |x_k| \left(\sum_{i=1}^m \sum_{k=1}^n |a_{ik}| \right) \\ = & \|A\|_{m_1}\|x\|_\infty, \end{aligned} \tag{5.3.1}$$

故范数$\|\cdot\|_1$相容于范数$\|\cdot\|_{m_1}$和$\|\cdot\|_\infty$. □

定理 5.3.2 $\|\cdot\|_{m_1}$和$\|\cdot\|_1$是相容的, 并且$\|\cdot\|_{m_1}$是$\mathbb{F}^{n\times n}$上的相容范数.

证明 由定理5.3.1得$\|\cdot\|_{m_1}$和$\|\cdot\|_1$是相容的.
任取$A, B \in \mathbb{F}^{n\times n}$, 令$B = [b_1\ b_2\ \cdots\ b_n]$, 则由式(5.3.1)得

$$\begin{aligned} \|AB\|_{m_1} = & \|A\,[b_1\ b_2\ \cdots\ b_n]\,\|_{m_1} \\ = & \sum_{j=1}^n \|Ab_j\|_1 \\ \leqslant & \sum_{j=1}^n \|A\|_{m_1}\|b_j\|_1 \\ = & \|A\|_{m_1} \sum_{j=1}^n \|b_j\|_1 \\ = & \|A\|_{m_1}\|B\|_{m_1}, \end{aligned}$$

故$\|\cdot\|_{m_1}$是$\mathbb{F}^{n\times n}$上的相容范数. □

定理 5.3.3 $\|\cdot\|_{m_\infty}$和$\|\cdot\|_\infty$是相容的, 并且$\|\cdot\|_{m_\infty}$是$\mathbb{F}^{n\times n}$上的相容范数.

证明 任取$A\in\mathbb{F}^{n\times n}, x\in\mathbb{F}^n$, 记$A=[a_{ij}], x=\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}^T$, 则

$$\begin{aligned}\|Ax\|_\infty &= \max_{1\leqslant i\leqslant n}\left|\sum_{k=1}^n a_{ik}x_k\right| \\ &\leqslant \max_{1\leqslant i\leqslant n}\left(\sum_{k=1}^n |a_{ik}||x_k|\right) \\ &\leqslant \left(\max_{1\leqslant i\leqslant n}\sum_{k=1}^n |a_{ik}|\right)\left(\max_k |x_k|\right) \\ &\leqslant \|A\|_{m_\infty}\|x\|_\infty,\end{aligned}$$

故$\|\cdot\|_{m_\infty}$和$\|\cdot\|_\infty$是相容的.

任取$A,B\in\mathbb{F}^{n\times n}$, 记$B=[b_1\ b_2\ \cdots\ b_n]$, 则由$\|\cdot\|_{m_\infty}$和$\|\cdot\|_\infty$相容得

$$\begin{aligned}\|AB\|_{m_\infty} &= \|[Ab_1\ Ab_2\ \cdots\ Ab_n]\|_{m_\infty} \\ &= \max_{1\leqslant j\leqslant n}\|Ab_j\|_\infty \\ &\leqslant \|A\|_{m_\infty}\|B\|_{m_\infty},\end{aligned}$$

故$\|\cdot\|_{m_\infty}$是$\mathbb{F}^{n\times n}$上的相容范数. □

注记 5.3.1 若删掉$\|\cdot\|_{m_\infty}$定义中的系数n, 则仍满足矩阵范数的定义, 但此时不是$\mathbb{F}^{n\times n}$上的相容范数. 例如: 当

$$A=\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix},\ B=\begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix}$$

时, 有$\|AB\|_{m_\infty}=2>1=\|A\|_{m_\infty}\|B\|_{m_\infty}$.

定理 5.3.4 $\|\cdot\|_{m_2}$和$\|\cdot\|_2$是相容的, 并且$\|\cdot\|_{m_2}$是$\mathbb{F}^{n\times n}$上的相容范数.

证明 任取$A\in\mathbb{F}^{n\times n}, x\in\mathbb{F}^n$, 记$A=[a_{ij}], x=\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}^T$, 则

由Hölder不等式推出

$$
\begin{aligned}
\|Ax\|_2^2 = & \sum_{i=1}^{n}\left|\sum_{k=1}^{n} a_{ik}x_k\right|^2 \\
\leqslant & \sum_{i=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}||x_k|\right)^2 \\
\leqslant & \sum_{i=1}^{n}\left[\left(\sum_{k=1}^{n}|a_{ik}|^2\right)\left(\sum_{k=1}^{n}|x_k|^2\right)\right] \\
= & \|x\|_2^2\left(\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}|^2\right) \\
= & \|A\|_{m_2}^2\cdot\|x\|_2^2,
\end{aligned}
$$

于是$\|Ax\|_2 \leqslant \|A\|_{m_2}\|x\|_2$, 故$\|\cdot\|_{m_2}$和$\|\cdot\|_2$是相容的.

任取$A,B\in\mathbb{F}^{n\times n}$, 记$B=[b_1\ b_2\ \cdots\ b_n]$, 则由$\|\cdot\|_{m_2}$和$\|\cdot\|_2$相容得

$$
\begin{aligned}
\|AB\|_{m_2}^2 = & \|A[b_1\ b_2\ \cdots\ b_n]\|_{m_2}^2 \\
= & \|[Ab_1\ Ab_2\ \cdots\ Ab_n]\|_{m_2}^2 \\
= & \sum_{j=1}^{n}\|Ab_j\|_2^2 \\
\leqslant & \|A\|_{m_2}^2\sum_{j=1}^{n}|b_j|_2^2 \\
= & \|A\|_{m_2}^2\|B\|_{m_2}^2,
\end{aligned}
$$

即$\|AB\|_{m_2}\leqslant\|A\|_{m_2}\|B\|_{m_2}$. 故$\|\cdot\|_{m_2}$是$\mathbb{F}^{n\times n}$上的相容范数. □

注记 5.3.2 通常也称$\|A\|_{m_2}$为Frobenius范数或$F-$范数.

5.3.2 从属范数

上节已证明了当$p=1,2,\infty$时, $\|\cdot\|_{m_p}$与$\|\cdot\|_p$是相容的. 那么, 对于$\mathbb{F}^n$上的任意范数$\|\cdot\|$, 可否找到$\mathbb{F}^{n\times n}$上的范数$\|\cdot\|_\alpha$使得$\|\cdot\|_\alpha$和$\|\cdot\|$相容? 为此, 定义$\mathbb{F}^{n\times n}$上的实值函数$\|\cdot\|_\alpha$如下:

$$
\|A\|_\alpha = \max_{\substack{x\in\mathbb{F}^n \\ \|x\|=1}}\|Ax\|,\ \forall A\in\mathbb{F}^{n\times n}. \tag{5.3.2}
$$

命题 5.3.1 设$\|\cdot\|$是$\mathbb{F}^n$上的范数, $\|\cdot\|_\alpha$由式(5.3.2)定义, 则

(i) $\|\cdot\|_\alpha$是$\mathbb{F}^{n\times n}$上的范数;

(ii) $\|\cdot\|_\alpha$与$\|\cdot\|$相容;

(iii) $\|\cdot\|_\alpha$是$\mathbb{F}^{n\times n}$上的相容范数.

证明 (i) 由式(5.3.2)易见$\|\cdot\|_\alpha$满足齐次性和三角不等式, 故只需证明$\|\cdot\|_\alpha$的非负性. 事实上, 显然有$\|A\|_\alpha \geqslant 0, \forall A \in \mathbb{F}^{n\times n}$. 进而

$$\begin{aligned} & \|A\|_\alpha = 0 \\ \Leftrightarrow\ & \|Ax\| = 0,\ \forall x \in \mathbb{F}^n \text{满足} \|x\| = 1 \\ \Leftrightarrow\ & Ax = 0,\ \forall x \in \mathbb{F}^n \text{满足} \|x\| = 1 \\ \Leftrightarrow\ & Ax = 0,\ \forall 0 \neq x \in \mathbb{F}^n \\ \Leftrightarrow\ & A = 0. \end{aligned}$$

(ii) 任取$A \in \mathbb{F}^{n\times n}$和非零的$x \in \mathbb{F}^n$, 易见

$$\|Ax\| = \|x\| \cdot \left\|A\frac{x}{\|x\|}\right\| \leqslant \|x\| \cdot \max_{\substack{y\in\mathbb{F}^n \\ \|y\|=1}} \|Ay\| = \|x\| \cdot \|A\|_\alpha,$$

于是$\|\cdot\|_\alpha$与$\|\cdot\|$相容.

(iii) 任取$A, B \in \mathbb{F}^{n\times n}$, 则由(ii)得

$$\|AB\|_\alpha = \max_{\substack{x\in\mathbb{F}^n \\ \|x\|=1}} \|ABx\| \leqslant \|A\|_\alpha \cdot \max_{\substack{x\in\mathbb{F}^n \\ \|x\|=1}} \|Bx\| = \|A\|_\alpha \cdot \|B\|_\alpha,$$

故$\|\cdot\|_\alpha$是$\mathbb{F}^{n\times n}$上的相容范数. □

定义 5.3.3 设$\|\cdot\|$是$\mathbb{F}^n$上的范数, $\|\cdot\|_\alpha$由式(5.3.2)定义, 称$\|\cdot\|_\alpha$是$\|\cdot\|$的从属范数.

命题 5.3.2 设$\|\cdot\|$是$\mathbb{F}^n$上的范数.

(i) 若$\|\cdot\|_\alpha$是$\|\cdot\|$的从属范数, 则$\|I_n\|_\alpha = 1$;

(ii) 若$\|\cdot\|_\beta$是$\mathbb{F}^{n\times n}$上与$\|\cdot\|$相容的范数, 则$\|I_n\|_\beta \geqslant 1$.

证明 (i) 由$\|\cdot\|_\alpha$是$\|\cdot\|$的从属范数知

$$\|I_n\|_\alpha = \max_{\substack{x\in\mathbb{F}^n \\ \|x\|=1}} \|I_n x\| = \max_{\substack{x\in\mathbb{F}^n \\ \|x\|=1}} \|x\| = 1.$$

(ii) 由$\|\cdot\|_\beta$与$\|\cdot\|$相容知

$$\|x\| = \|I_n x\| \leqslant \|I_n\|_\beta \|x\|,\ \forall\, x \in \mathbb{F}^n,$$

故$\|I_n\|_\beta \geqslant 1$. □

对于$\mathbb{F}^n$上范数$\|\cdot\|$, 习惯上仍将其从属范数记作$\|\cdot\|$. 下面进一步给出$\mathbb{F}^n$上的范数$\|\cdot\|_p (p = 1, 2, \infty)$的从属范数的精细刻画.

定理 5.3.5 $\|A\|_2=\sqrt{\lambda_{\max}(A^*A)}=\sqrt{\lambda_{\max}(AA^*)},\ \forall A\in\mathbb{F}^{n\times n}$.

证明 任取$x\in\mathbb{F}^n$, 则$\|x\|_2=\sqrt{x^*x}$, 于是$\|Ax\|_2^2=x^*A^*Ax$, 故由定理2.4.4得

$$\|A\|_2=\max_{\substack{x\in\mathbb{F}^n\\ \|x\|_2=1}}\|Ax\|_2=\max_{\substack{x\in\mathbb{F}^n\\ \|x\|_2=1}}\sqrt{x^*A^*Ax}=\sqrt{\lambda_{\max}(A^*A)}.$$

再由AB与BA有相同的非零特征值得证. □

定理 5.3.6 $\|A\|_1=\max\limits_{1\leqslant j\leqslant n}\sum\limits_{i=1}^n|a_{ij}|,\quad \forall A=[a_{ij}]\in\mathbb{F}^{n\times n}$.

证明 任取$x\in\mathbb{F}^n$, 则$\|x\|_1=\sum\limits_{i=1}^n|x_i|$, 于是

$$\begin{aligned}\|Ax\|_1 &= \sum_{i=1}^n\left|\sum_{j=1}^n a_{ij}x_j\right|\\ &\leqslant \sum_{i=1}^n\sum_{j=1}^n|a_{ij}||x_j|\\ &= \sum_{j=1}^n\left(|x_j|\sum_{i=1}^n|a_{ij}|\right)\\ &\leqslant \sum_{j=1}^n|x_j|\left(\max_{1\leqslant j\leqslant n}\sum_{i=1}^n|a_{ij}|\right)\\ &= \|x\|_1\cdot\max_{1\leqslant j\leqslant n}\sum_{i=1}^n|a_{ij}|,\end{aligned}$$

故

$$\|A\|_1=\max_{\substack{x\in\mathbb{F}^n\\ \|x\|_1=1}}\|Ax\|_1\leqslant\max_{1\leqslant j\leqslant n}\sum_{i=1}^n|a_{ij}|.$$

现在只需证上式中的等号能够取到. 事实上, 设

$$\max_{1\leqslant j\leqslant n}\sum_{i=1}^n|a_{ij}|=\sum_{i=1}^n|a_{ij_0}|,$$

取$x_0=e_{j_0}$, 则$\|x_0\|_1=\|e_{j_0}\|_1=1$且$\|Ax_0\|_1=\|Ae_{j_0}\|_1=\sum\limits_{j=1}^n|a_{ij_0}|$. 总之, $\|A\|_1=\max\limits_{1\leqslant j\leqslant n}\sum\limits_{i=1}^n|a_{ij}|$. □

定理 5.3.7 $\|A\|_\infty=\max\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^n|a_{ij}|,\ \forall A=[a_{ij}]\in\mathbb{F}^{n\times n}$.

证明 任取$x \in \mathbb{F}^n$, 则$\|x\|_\infty = \max\limits_{1\leqslant i\leqslant n} |x_i|$, 于是

$$\begin{aligned}\|Ax\|_\infty &= \max_{1\leqslant i\leqslant n}\left|\sum_{j=1}^n a_{ij}x_j\right| \\ &\leqslant \max_{1\leqslant i\leqslant n}\sum_{j=1}^n |a_{ij}||x_j| \\ &\leqslant \left(\max_{1\leqslant j\leqslant n}|x_j|\right)\left(\max_{1\leqslant i\leqslant n}\sum_{j=1}^n |a_{ij}|\right) \\ &= \|x\|_\infty \max_{1\leqslant i\leqslant n}\sum_{j=1}^n |a_{ij}|,\end{aligned}$$

故

$$\|A\|_\infty = \max_{\substack{x\in\mathbb{F}^n \\ \|x\|_\infty=1}} \|Ax\|_\infty \leqslant \max_{1\leqslant i\leqslant n}\sum_{j=1}^n |a_{ij}|.$$

现在只需证上式中的等号能够取到. 事实上, 设$\max\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^n |a_{ij}| = \sum\limits_{j=1}^n |a_{i_0j}|$, 取$x_0 = \begin{bmatrix} x_1 & x_2 & \cdots & x_n\end{bmatrix}^T$满足

$$x_j = \begin{cases} \dfrac{a_{i_0j}}{|a_{i_0j}|}, & a_{i_0j} \neq 0, \\ 1, & a_{i_0j} = 0, \end{cases} \qquad j = 1, 2, \cdots, n,$$

则$\|x_0\|_\infty = 1$且

$$\|Ax_0\|_\infty = \max_{1\leqslant i\leqslant n}\left|\sum_{j=1}^n a_{ij}x_j\right| = \sum_{j=1}^n |a_{i_0j}|.$$

因而$\|A\|_\infty = \max\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^n |a_{ij}|$. □

基于上面的三个定理, 也称$\mathbb{F}^{n\times n}$上的$\|\cdot\|_2$范数为谱范数、$\|\cdot\|_1$范数为列和范数、$\|\cdot\|_\infty$范数为行和范数.

可模仿$\mathbb{F}^{n\times n}$上的谱范数、行和范数、列和范数, 分别定义$\mathbb{F}^{m\times n}$上的范数如下:

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^*A)},\ \forall A \in \mathbb{F}^{m\times n};$$

$$\|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|,\ \forall A = [a_{ij}] \in \mathbb{F}^{m\times n};$$

$$\|A\|_1 = \max_j \sum_{i=1}^m |a_{ij}|,\ \forall A = [a_{ij}] \in \mathbb{F}^{m\times n}.$$

5.4 谱半径和条件数

定义 5.4.1 设$A \in \mathbb{F}^{n\times n}$有特征根$\lambda_1, \lambda_2, \cdots, \lambda_n$, 称$\max\limits_{1\leqslant i\leqslant n} |\lambda_i|$为$A$的谱半径, 记作$\rho(A)$.

定理 5.4.1 设$\|\cdot\|$是$\mathbb{F}^{n\times n}$上的从属范数, 则

$$\rho(A) \leqslant \|A\|, \ \forall\ A \in \mathbb{F}^{n\times n}.$$

证明 任取A的特征根λ, 设x是A的属于λ的特征向量, 则$Ax = \lambda x$. 又因$\|\cdot\|$是$\mathbb{F}^{n\times n}$上的从属范数, 故

$$|\lambda|\|x\| = \|\lambda x\| = \|Ax\| \leqslant \|A\|\|x\|,$$

于是$|\lambda| \leqslant \|A\|$. 再由谱半径定义得$\rho(A) \leqslant \|A\|$. □

定义 5.4.2 设$\|\cdot\|$是$\mathbb{F}^{n\times n}$上的范数, $A \in GL_n(\mathbb{F})$, 称$\|A\|\cdot\|A^{-1}\|$是A的条件数, 记作$\mathrm{cond}_{\|\cdot\|}(A)$, 简记为$\mathrm{cond}(A)$.

性质 5.4.1 设$\|\cdot\|$是$\mathbb{F}^{n\times n}$上的相容矩阵范数, $A \in \mathbb{F}^{n\times n}$可逆, 则

(i) $\mathrm{cond}_{\|\cdot\|}(A) \geqslant 1$;

(ii) $\mathrm{cond}_{\|\cdot\|}(kA) = \mathrm{cond}_{\|\cdot\|}(A), \ \forall\ k \in \mathbb{F}$;

(iii) 若A为酉阵, 则$\mathrm{cond}_{\|\cdot\|_2}(A) = 1$;

(iv) 设U和V是酉阵, 则$\mathrm{cond}_{\|\cdot\|_2}(UA) = \mathrm{cond}_{\|\cdot\|_2}(AV) = \mathrm{cond}_{\|\cdot\|_2}(A)$.

证明 (i) 由条件数及相容范数的定义和命题5.3.2知

$$\mathrm{cond}_{\|\cdot\|}(A) = \|A\| \cdot \|A^{-1}\| \geqslant \|AA^{-1}\| = \|I_n\| \geqslant 1.$$

(ii) 由条件数及范数定义知

$$\mathrm{cond}_{\|\cdot\|}(kA) = \|kA\| \cdot \|(kA)^{-1}\| = \|A\| \cdot \|A^{-1}\| = \mathrm{cond}_{\|\cdot\|}(A).$$

(iii) 由A为酉阵知$\|A\|_2 = \sqrt{\lambda_{\max}(A^*A)} = 1$, 故

$$\mathrm{cond}_{\|\cdot\|_2}(A) = \|A\|_2 \cdot \|A^{-1}\|_2 = 1.$$

(iv) 易见

$$\|UA\|_2 = \sqrt{\lambda_{\max}\left((UA)^*(UA)\right)} = \sqrt{\lambda_{\max}(A^*A)} = \|A\|_2.$$

同理

$$\|(UA)^{-1}\|_2 = \|A^{-1}\|_2.$$

故

$$\mathrm{cond}_{\|\cdot\|_2}(UA) = \|UA\|_2 \cdot \|(UA)^{-1}\|_2 = \|A\|_2 \cdot \|A^{-1}\|_2 = \mathrm{cond}_{\|\cdot\|_2}(A).$$

同理$\mathrm{cond}_{\|\cdot\|_2}(AV) = \mathrm{cond}_{\|\cdot\|_2}(A)$. □

5.5 矩阵测度

本节总是假设$\|\cdot\|$是$\mathbb{F}^{n\times n}$上的从属范数.

引理 5.5.1 若$A \in \mathbb{F}^{n\times n}$, $h > 0$, 则

$$-\|A\| \leqslant -\frac{\|I_n - hA\| - 1}{h} \leqslant \frac{\|I_n + hA\| - 1}{h} \leqslant \|A\|.$$

证明 由$\dfrac{\|I_n - hA\| - 1}{h} + \dfrac{\|I_n + hA\| - 1}{h} \geqslant \dfrac{\|2I_n\| - 2}{h} = 0$得

$$\begin{aligned}-\|A\| = & -\frac{1 + h\|A\| - 1}{h} \leqslant -\frac{\|I_n - hA\| - 1}{h} \leqslant \frac{\|I_n + hA\| - 1}{h} \\ \leqslant & \ \frac{1 + h\|A\| - 1}{h} = \|A\|.\end{aligned}$$

□

定理 5.5.1 设$A \in \mathbb{F}^{n\times n}$, 则$\lim\limits_{h\to 0^+} \dfrac{\|I_n + hA\| - 1}{h}$存在.

证明 显然$\dfrac{\|I_n + hA\| - 1}{h}$在$(0, +\infty)$上连续. 对于满足$0 < \alpha < 1$的实数$\alpha$, 有

$$\frac{\|I_n + \alpha hA\| - 1}{\alpha h} = \frac{\|\alpha(I_n + hA) + (1-\alpha)I_n\| - 1}{\alpha h} \leqslant \frac{\|I_n + hA\| - 1}{h},$$

于是$\dfrac{\|I_n + hA\| - 1}{h}$关于$h$单调递减. 由引理5.5.1知$\dfrac{\|I_n + hA\| - 1}{h}$有界, 故

$$\lim_{h\to 0^+} \frac{\|I_n + hA\| - 1}{h}$$

存在. □

定义 5.5.1 称极限$\lim\limits_{h\to0^+}\dfrac{\|I_n+hA\|-1}{h}$为矩阵$A$关于范数$\|\cdot\|$的测度，记作$\mu_{\|\cdot\|}(A)$. 当不强调范数的类型时，简记为$\mu(A)$.

下面的定理给出了矩阵测度的一些性质.

定理 5.5.2 设$A, B\in\mathbb{F}^{n\times n}$, 则下面的(i)~(v)成立.

(i) $-\|A\|\leqslant-\mu_{\|\cdot\|}(-A)\leqslant\mu_{\|\cdot\|}(A)\leqslant\|A\|$;

(ii) $\mu(0)=0,\ \mu(I_n)=1,\ \mu(-I_n)=-1$;

(iii) $\mu(cA)=\begin{cases}c\mu(A), & 若c\geqslant0,\\ \mu(cA)=-c\mu(-A), & 若c<0;\end{cases}$

(iv) $\max\{\mu(A)-\mu(-B),-\mu(-A)+\mu(B)\}\leqslant\mu(A+B)\leqslant\mu(A)+\mu(B)$;

(v) $\mu(\cdot)$是$\mathbb{F}^{n\times n}$上的凸函数，即$\mu[\alpha A+(1-\alpha)B]\leqslant\alpha\mu(A)+(1-\alpha)\mu(B)$, $\forall\alpha\in[0,1]$.

证明 (i) 在引理5.5.1中令$h\to0^+$即得.

(ii) 由定义5.5.1易见.

(iii) 当$c=0$时，由(ii)得到. 当$c>0$时，由定义5.5.1知

$$\mu(cA)=\lim_{h\to0^+}\frac{\|I_n+h(cA)\|-1}{h}=\lim_{h\to0^+}\frac{\|I_n+(ch)A\|-1}{ch}\cdot c=c\mu(A).$$

当$c<0$时，由定义5.5.1知

$$\begin{aligned}\mu(cA)=&\lim_{h\to0^+}\frac{\|I_n+h(cA)\|-1}{h}\\=&\lim_{h\to0^+}\frac{\|I_n+(-ch)(-A)\|-1}{-ch}\cdot(-c)\\=&-c\mu(-A).\end{aligned}$$

(iv) 由矩阵测度的定义和范数的性质知

$$\begin{aligned}\mu(A+B) &= \lim_{h\to 0^+}\frac{\|I_n+h(A+B)\|-1}{h}\\ &= \lim_{h\to 0^+}\frac{\left\|\frac{1}{2}I_n+hA+\frac{1}{2}I_n+hB)\right\|-1}{h}\\ &\leqslant \lim_{h\to 0^+}\frac{\left(\left\|\frac{1}{2}I_n+hA\right\|-\frac{1}{2}+\left\|\frac{1}{2}I_n+hB\right\|-\frac{1}{2}\right)}{h}\\ &= \lim_{h\to 0^+}\left[\frac{\|I_n+(2h)A\|-1}{2h}+\frac{\|I_n+(2h)B\|-1}{2h}\right]\\ &= \mu(A)+\mu(B).\end{aligned}$$

又$\mu(A)=\mu(A+B-B)\leqslant\mu(A+B)+\mu(-B)$, 则

$$\mu(A)-\mu(-B)\leqslant\mu(A+B).$$

同理

$$-\mu(-A)+\mu(B)\leqslant\mu(A+B).$$

故$\max\{\mu(A)-\mu(-B),-\mu(-A)+\mu(B)\}\leqslant\mu(A+B)$.

(v) 组合(iii)和(iv)得到. □

定理 5.5.3 设$A\in\mathbb{F}^{n\times n}$, λ是A的特征根, 则

$$-\mu_{\|\cdot\|}(-A)\leqslant\mathrm{Re}\lambda\leqslant\mu_{\|\cdot\|}(A).$$

证明 设$\|\cdot\|$为向量范数$\|\cdot\|_v$的从属范数, z是A的属于λ的单位特征向量, 于是$(I_n+hA)z=(1+h\lambda)z$, 因而

$$\|I_n+hA\|=\max_{\|x\|_v=1}\|(I_n+hA)x\|_v\geqslant\|(I_n+hA)z\|_v=|1+h\lambda|,$$

故

$$\frac{\|I_n+hA\|-1}{h}\geqslant\frac{|1+h\lambda|-1}{h}.$$

当$h\to 0^+$时, 上式左端收敛于$\mu_{\|\cdot\|}(A)$, 而右端收敛于$\mathrm{Re}\lambda$, 从而$\mu_{\|\cdot\|}(A)\geqslant\mathrm{Re}\lambda$. 又$\mu_{\|\cdot\|}(-A)\geqslant\mathrm{Re}(-\lambda)=-\mathrm{Re}\lambda$, 即$-\mu_{\|\cdot\|}(-A)\leqslant\mathrm{Re}\lambda$. □

关于矩阵范数$\|\cdot\|_p(p=1,2,\infty)$的矩阵测度$\mu_{\|\cdot\|_p}$的具体刻画见下列命题(证明见附录D).

命题 5.5.1 设$A\in\mathbb{F}^{n\times n}$, 则$\mu_{\|\cdot\|_2}(A)=\frac{1}{2}\lambda_{\max}(A+A^*)$.

命题 5.5.2 设$A \in \mathbb{F}^{n\times n}$, 则$\mu_{\|\cdot\|_1}(A) = \max\limits_{1\leqslant j\leqslant n}\{\mathrm{Re}(a_{jj}) + \sum\limits_{\substack{i=1\\i\neq j}}^{n}|a_{ij}|\}$.

命题 5.5.3 设$A \in \mathbb{F}^{n\times n}$, 则$\mu_{\|\cdot\|_\infty}(A) = \max\limits_{1\leqslant i\leqslant n}\{\mathrm{Re}(a_{ii}) + \sum\limits_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|\}$.

由定理5.5.3可知为了估计一个已知矩阵特征值实部的范围, 可以利用不同的测度函数来求出实轴上的不同估计区间. 因此, 为了求得最好的估计, 我们希望取所有区间的交集.

例 5.5.1 设$A = \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix}$, 则$A$的特征值是2和1, 所以包含$A$的两个特征值的实部的最小区间是$[1,2]$. 现在用测度函数$\mu_{\|\cdot\|_1}$, $\mu_{\|\cdot\|_2}$和$\mu_{\|\cdot\|_\infty}$来估计这个区间. 用$\mu_{\|\cdot\|_1}$估计得$\mathrm{Re}\lambda(A) \in [0,2]$, 用$\mu_{\|\cdot\|_\infty}$估计得$\mathrm{Re}\lambda(A) \in [1,3]$, 用$\mu_{\|\cdot\|_2}$估计得$\mathrm{Re}\lambda(A) \in [\frac{3}{2} - \frac{1}{\sqrt{2}}, \frac{3}{2} + \frac{1}{\sqrt{2}}] \approx [0.793, 2.207]$, 因此, $\mu_{\|\cdot\|_1}$给出一个准确的上界, $\mu_{\|\cdot\|_\infty}$给出一个准确的下界, 而$\mu_{\|\cdot\|_2}$给出一个最小"宽度"的区间. 易见三个区间的交为$[1,2]$.

由例5.5.1可知, 测度函数在求矩阵特征值实部的较准确的上下界时是很有用的, 从而可用来判断系统的稳定性.

习 题

1. 证明: $|\|x\| - \|y\|| \leqslant \|x-y\|$, $\forall x, y \in \mathbb{F}^n$.

2. 证明: $\frac{1}{2}\left(\|x+y\|_2^2 + \|x-y\|_2^2\right) = \|x\|_2^2 + \|y\|_2^2$, $\forall x, y \in \mathbb{F}^n$.

3. 设$\|\cdot\|_a$是$\mathbb{C}^n$的一个向量范数, 矩阵$A \in \mathbb{C}^{m\times n}$是列满秩的, 证明:

$$\|X\|_A = \|AX\|_a, \ \forall X \in \mathbb{C}^n$$

   所定义的函数是$\mathbb{C}^n$的一个向量范数.

4. 设$A, B \in \mathbb{C}^{n\times n}$可逆, 对$\mathbb{C}^n$中任意向量$\alpha$, 定义

$$\|\alpha\| = \|A\alpha\|_1 + 3\|B\alpha\|_2, \ \forall \alpha \in \mathbb{C}^n.$$

   验证: $\|\alpha\|$是$\mathbb{C}^n$中的向量范数.

5. 证明: 方阵的m_1-范数与F$-$范数等价.

6. 证明: 方阵的$m_\infty-$范数与向量1$-$范数相容.

7. 定义$\|A\|_G = \sqrt{mn}\max\limits_{i,j}|a_{ij}|$, $\forall A \in \mathbb{C}^{m\times n}$, 证明: $\|A\|_G$是$\mathbb{C}^{m\times n}$中的矩阵范数, 且与向量2-范数相容.

8. 证明: $\mathbb{C}^{n\times n}$中矩阵范数$\|\cdot\|_{m_1}$与$\mathbb{C}^n$中向量范数$\|\cdot\|_p$ $(p \geqslant 1)$相容.

9. 设

$$A = \begin{bmatrix} -1 & 0 & 2 & i \\ 3+i & 5 & 1+i & 0 \\ 2 & i & 2 & -4 \end{bmatrix},$$

$$X = \begin{bmatrix} -1 & 2 & 0 & -i \end{bmatrix}^T,$$

计算$\|AX\|_1$, $\|AX\|_2$, $\|AX\|_\infty$, $\|A\|_1$, $\|A\|_\infty$.

10. 设$A \in \mathbb{F}^{n\times n}$, 则

(i) $\|A\|_2 = \max\limits_{\substack{\|x\|_2=1 \\ \|y\|_2=1}} |y^*Ax|$;

(ii) $\|A^*\|_2 = \|A\|_2$;

(iii) $\|A^*A\|_2 = (\|A\|_2)^2$.

(提示: 使用$\|A\|_2 = \max\limits_{\|x\|_2=1}\sqrt{x^*A^*Ax} = \max\limits_{\|x\|_2=1}\left|(\frac{Ax}{\|Ax\|_2})^*Ax\right|$.)

11. 设$A \in \mathbb{F}^{n\times n}$, U和V是酉矩阵, 则

(i) $\|UAV\|_F = \|A\|_F$;

(ii) 当A可逆时有$\mathrm{cond}_{\|\cdot\|_F}(AV) = \mathrm{cond}_{\|\cdot\|_F}(UA) = \mathrm{cond}_{\|\cdot\|_F}(A)$.

12. 设$\|\cdot\|$是$\mathbb{C}^{n\times n}$上的任一相容从属范数, 则对任意$A \in \mathbb{C}^{n\times n}$有

$$|\lambda_i| \leqslant \|A\|,\ \forall\ \lambda_i \in \lambda(A).$$

13. 已知

$$A = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{6} & \frac{1}{6} & \frac{3}{6} & \frac{1}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{3}{7} \end{bmatrix},$$

证明: 矩阵A的谱半径$\rho(A) \leqslant 1$.

14. 设$A = [a_{ij}] \in \mathbb{C}^{n\times n}$, $\lambda_1, \lambda_2, \cdots, \lambda_n$是$A$的特征值, 证明: $\sum\limits_{i=1}^n \lambda_i^2 \leqslant \sum\limits_{i=1}^n\sum\limits_{j=1}^n |a_{ij}|^2$, 并且等号成立的充要条件是$A$为正规矩阵.

(提示: 使用$\|A\|_F^2 = \mathrm{tr}(AA^*)$, 并且注意$\|\cdot\|_F$是酉不变范数.)

15. 证明: $\mu(A+\alpha I_n)=\mu(A)+\alpha,\ \forall\alpha\in\mathbb{R},A\in\mathbb{F}^{n\times n}$.

16. 证明: $|\mu(A)-\mu(B)|\leqslant\max\{\mu(A-B),\mu(B-A)\},\ \forall A,B\in\mathbb{F}^{n\times n}$.

第6章　矩阵序列的极限与矩阵级数

本章将数列的极限、级数、幂级数的概念及其性质推广到矩阵序列的极限、矩阵级数和矩阵幂级数.

6.1　矩阵序列的极限

定义 6.1.1　设$A^{(k)}=[a_{ij}^{(k)}]\in\mathbb{R}^{m\times n}$, $k=1,2,\cdots,A=[a_{ij}]\in\mathbb{R}^{m\times n}$. 若

$$\lim_{k\to+\infty}a_{ij}^{(k)}=a_{ij},\ \forall\ i,j,$$

称A为序列$\{A^{(k)}\}$当$k\to+\infty$时的极限, 记作

$$\lim_{k\to+\infty}A^{(k)}=A\text{ 或 }A^{(k)}\to A\ (k\to+\infty).$$

此时, 称$\{A^{(k)}\}$是收敛的, 否则称$\{A^{(k)}\}$是发散的.

定义 6.1.2　设$A\in\mathbb{R}^{m\times n}$, $\{A^{(k)}\}$为$\mathbb{R}^{m\times n}$中的矩阵序列, $\|\cdot\|$是$\mathbb{R}^{m\times n}$上的范数, 若

$$\lim_{k\to+\infty}\|A^{(k)}-A\|=0,$$

称矩阵序列$\{A^{(k)}\}$依范数$\|\cdot\|$收敛于A.

定理 6.1.1　设$A\in\mathbb{R}^{m\times n}$, $\{A^{(k)}\}$为$\mathbb{R}^{m\times n}$中矩阵序列, $\|\cdot\|$是$\mathbb{R}^{m\times n}$上的范数, 则$\lim\limits_{k\to+\infty}A^{(k)}=A$的充要条件是

$$\lim_{k\to+\infty}\|A^{(k)}-A\|=0.\tag{6.1.1}$$

证明 记$A^{(k)}=[a_{ij}^{(k)}], A=[a_{ij}]$, 则

$$\begin{aligned}\lim_{k\to+\infty}A^{(k)}=A \Leftrightarrow\ & \lim_{k\to+\infty}a_{ij}^{(k)}=a_{ij},\ \forall\ i,j\\ \Leftrightarrow\ & \lim_{k\to+\infty}|a_{ij}^{(k)}-a_{ij}|=0,\ \forall\ i,j\\ \Leftrightarrow\ & \lim_{k\to+\infty}\sum_{j=1}^{n}|a_{ij}^{(k)}-a_{ij}|=0,\ \forall\ i\\ \Leftrightarrow\ & \lim_{k\to+\infty}\max_{1\leqslant i\leqslant m}\sum_{j=1}^{n}|a_{ij}^{(k)}-a_{ij}|=0.\\ \Leftrightarrow\ & \lim_{k\to+\infty}\|A^{(k)}-A\|_\infty=0.\\ \Leftrightarrow\ & \lim_{k\to+\infty}\|A^{(k)}-A\|=0.\end{aligned}$$

□

由范数的等价性易见: 若$\mathbb{R}^{m\times n}$中的矩阵序列$\{A^{(k)}\}$依某种范数收敛, 则$\{A^{(k)}\}$依任意范数都收敛.

推论 6.1.1 设$A\in\mathbb{R}^{m\times n}$, $\{A^{(k)}\}$为$\mathbb{R}^{m\times n}$中的矩阵序列, $\|\cdot\|$是$\mathbb{R}^{m\times n}$上的范数. 若$\lim\limits_{k\to+\infty}A^{(k)}=A$, 则$\lim\limits_{k\to+\infty}\|A^{(k)}\|=\|A\|$.

证明 由定理6.1.1及$|\ \|A^{(k)}\|-\|A\|\ |\leqslant\|A^{(k)}-A\|$有$\lim\limits_{k\to\infty}|\ \|A^{(k)}-\|A\|\ |=0$, 即$\lim\limits_{k\to+\infty}\|A^{(k)}\|=\|A\|$. □

性质 6.1.1 设$\mathbb{R}^{m\times n}$中的矩阵序列$\{A^{(k)}\}$和$\{B^{(k)}\}$满足

$$\lim_{k\to+\infty}A^{(k)}=A,\ \lim_{k\to+\infty}B^{(k)}=B,$$

则

(i) $\lim\limits_{k\to+\infty}\left(A^{(k)}\pm B^{(k)}\right)=A\pm B$;

(ii) $\lim\limits_{k\to+\infty}\left(xA^{(k)}\right)=xA, \forall\ x\in\mathbb{R}$;

(iii) $\lim\limits_{k\to+\infty}\left(A^{(k)}B^{(k)}\right)=AB$;

(iv) 当$P\in\mathbb{R}^{m\times m}, Q\in\mathbb{R}^{n\times n}$时, $\lim\limits_{k\to+\infty}\left(PA^{(k)}Q\right)=PAQ$;

(v) 当A和$A^{(k)}$可逆, $k=1,2,\cdots$, $\lim\limits_{k\to+\infty}\left(A^{(k)}\right)^{-1}=A^{-1}$.

证明 (i)~(iv)显然成立, 下证(v). 由(iii)知

$$A\cdot \lim_{k\to+\infty}\left(A^{(k)}\right)^{-1}=\lim_{k\to+\infty}\left(A^{(k)}\cdot\left(A^{(k)}\right)^{-1}\right)=\lim_{k\to+\infty}I_n=I_n,$$

由逆矩阵的唯一性知 $\lim_{k\to+\infty}\left(A^{(k)}\right)^{-1}=A^{-1}$. 故(v)成立. □

定理 6.1.2 $\lim_{k\to+\infty}A^k=0 \Leftrightarrow \rho(A)<1$.

证明 设$A=P\mathrm{diag}(J_1,J_2,\cdots,J_s)P^{-1}$, 其中

$$J_i=\begin{bmatrix}\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{bmatrix},\quad i=1,\cdots,s,$$

则$A^k=P\mathrm{diag}(J_1^k,J_2^k,\cdots,J_s^k)P^{-1}$, 故

$$\begin{aligned}\lim_{k\to+\infty}A^k=0 &\Leftrightarrow \lim_{k\to+\infty}J_i^k=0,\ i=1,2,\cdots,s\\ &\Leftrightarrow \lim_{k\to+\infty}\lambda_i^k=0,\ i=1,2,\cdots,s\\ &\Leftrightarrow |\lambda_i|<1, i=1,2,\cdots,s\\ &\Leftrightarrow \rho(A)<1.\end{aligned}$$

□

推论 6.1.2 设$\|\cdot\|$是从属范数, 若A满足$\|A\|<1$, 则$\lim_{k\to+\infty}A^k=0$.

证明 由定理5.4.1知$\rho(A)\leqslant\|A\|$. 又$\|A\|<1$, 故$\rho(A)<1$. 由定理6.1.2得证. □

6.2 矩阵级数

对于矩阵序列$A^{(0)}$, $A^{(1)}$, $A^{(2)}$, $\cdots$, 令$S^{(k)}=\sum_{i=0}^{k}A^{(i)}$, $k=0,1,2,\cdots$, 则$S^{(0)}$, $S^{(1)}$, $S^{(2)}$, $\cdots$也是矩阵序列.

定义 6.2.1 若矩阵序列$S^{(k)}$收敛, 称矩阵级数$\sum_{k=0}^{+\infty}A^{(k)}$收敛, 否则称矩阵级数$\sum_{k=0}^{+\infty}A^{(k)}$发散. 若$\lim_{k\to+\infty}S^{(k)}=S$, 称$S$是矩阵级数$\sum_{k=0}^{+\infty}A^{(k)}$的和, 记作$\sum_{k=0}^{+\infty}A^{(k)}=S$. 若级数$\sum_{k=0}^{+\infty}|a_{ij}^{(k)}|$ $(\forall\, i,j)$收敛, 称$\sum_{k=0}^{+\infty}A^{(k)}$绝对收敛.

类似于数列级数, 矩阵级数有下面的性质.

性质 6.2.1 设$\{A^{(k)}\}$和$\{B^{(k)}\}$是$\mathbb{R}^{m\times n}$中的矩阵序列, $T, S=[s_{ij}]\in\mathbb{R}^{m\times n}$, 则下面的命题(i)~(iv)成立:

(i) $\sum\limits_{k=0}^{+\infty}A^{(k)}=S$的充要条件是$\sum\limits_{k=0}^{+\infty}a_{ij}^{(k)}=s_{ij},\ \forall\ i,j$;

(ii) 若$\sum\limits_{k=0}^{+\infty}A^{(k)}$收敛, 则$\lim\limits_{k\to+\infty}A^{(k)}=0$;

(iii) 若$\sum\limits_{k=0}^{+\infty}A^{(k)}=S$, $\sum\limits_{k=0}^{+\infty}B^{(k)}=T$, 则$\sum\limits_{k=0}^{+\infty}\left(A^{(k)}\pm B^{(k)}\right)=S\pm T$;

(iv) 若$\sum\limits_{k=0}^{+\infty}A^{(k)}=S$, 则$\sum\limits_{k=0}^{+\infty}\left(\mu A^{(k)}\right)=\mu S,\ \forall\ \mu\in\mathbb{R}$.

定理 6.2.1 矩阵级数$\sum\limits_{k=0}^{+\infty}A^{(k)}$绝对收敛的充要条件是级数$\sum\limits_{k=0}^{+\infty}\|A^{(k)}\|$对任意(某种)范数$\|\cdot\|$收敛.

证明

$$
\begin{aligned}
\sum_{k=0}^{+\infty}A^{(k)}\text{绝对收敛}\ &\Leftrightarrow \sum_{k=0}^{+\infty}|a_{ij}^{(k)}|\text{收敛},\ \forall\ i,j\\
&\Leftrightarrow \sum_{k=0}^{+\infty}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}^{(k)}|\right)\text{收敛}\\
&\Leftrightarrow \sum_{k=0}^{+\infty}\|A^{(k)}\|_{m_1}\text{收敛}\\
&\Leftrightarrow \sum_{k=0}^{+\infty}\|A^{(k)}\|\text{对任意(某种)范数}\|\cdot\|\text{收敛}.
\end{aligned}
$$

□

注意到

$$\sum_{k=0}^{+\infty}\frac{\|A^k\|}{k!}\leqslant\sum_{k=0}^{+\infty}\frac{\|A\|^k}{k!}=\mathrm{e}^{\|A\|},$$

所以$\sum\limits_{k=0}^{+\infty}\frac{A^k}{k!}$绝对收敛. 定义$\mathrm{e}^A=\sum\limits_{k=0}^{+\infty}\frac{A^k}{k!}$. 类似地, 按照$\sin x$和$\cos x$在$x=0$处的Taylor展开式可以定义$\sin A$和$\cos A$如下:

$$\sin A=\sum_{k=0}^{+\infty}\frac{(-1)^k A^{2k+1}}{(2k+1)!},\ \cos A=\sum_{k=0}^{+\infty}\frac{(-1)^k A^{2k}}{(2k)!},\ \forall\ A\in\mathbb{F}^{n\times n}.$$

6.3 矩阵幂级数

定义 6.3.1 设$A\in\mathbb{R}^{m\times n}$, $c_k\in\mathbb{R}$, $k=1,2,\cdots$, 称$\sum\limits_{k=0}^{+\infty}c_kA^k$为幂级数.

定理 6.3.1 设$\|\cdot\|$是$\mathbb{R}^{m\times n}$上的范数, 若正项级数$\sum\limits_{k=0}^{+\infty}|c_k|\cdot\|A\|^k$收敛, 则幂级数$\sum\limits_{k=0}^{+\infty}c_kA^k$绝对收敛.

证明 由$\sum\limits_{k=0}^{+\infty}|c_k|\cdot\|A\|^k$收敛及

$$\sum_{k=0}^{+\infty}\|c_kA^k\|\leqslant\sum_{k=0}^{+\infty}|c_k|\cdot\|A\|^k$$

知$\sum\limits_{k=0}^{+\infty}\|c_kA^k\|$收敛, 从而$\sum\limits_{k=0}^{+\infty}c_kA^k$绝对收敛. □

定理 6.3.2 若A的谱半径落在$\sum\limits_{k=0}^{+\infty}c_kz^k$的收敛域内, 则$\sum\limits_{k=0}^{+\infty}c_kA^k$收敛.

证明 设$f(z)=\sum\limits_{k=0}^{+\infty}c_kz^k$是收敛半径为$r$的幂级数, 则$\rho(A)<r$. 设

$$A=P\mathrm{diag}(J_1,J_2,\cdots,J_t)P^{-1},$$

其中

$$J_i=\begin{bmatrix}\lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1\\ & & & & \lambda_i\end{bmatrix}\in\mathbb{R}^{n_i\times n_i},\ i=1,2,\cdots,t,$$

则

$$f(A)=P\mathrm{diag}\left(f(J_1),f(J_2),\cdots,f(J_t)\right)P^{-1}.$$

下面考虑$f(J_i)$的收敛性. 注意到

$$J_i^k=\left(\lambda_iI+\begin{bmatrix}0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & 0\end{bmatrix}\right)^k=\begin{bmatrix}b_0 & b_1 & \cdots & b_{n_i-1}\\ & \ddots & \ddots & \vdots\\ & & \ddots & b_1\\ & & & b_0\end{bmatrix},$$

其中$b_j = C_k^j \lambda_i^{k-j}, j = 0, 1, \cdots, n_i - 1, C_k^p$表示从$k$个元素中任取$p$个的组合数，但规定当$p > k$时$C_k^p = 0$. 故

$$f(J_i) = \sum_{k=0}^{+\infty} c_k J_i^k = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n_i-1} \\ & \ddots & \ddots & \vdots \\ & & \ddots & a_1 \\ & & & a_0 \end{bmatrix},$$

其中$a_j = \sum\limits_{k=j}^{+\infty} c_k b_k$. 注意到

$$a_j = \sum_{k=j}^{+\infty} \frac{c_k \cdot k!}{j!(k-j)!} \lambda_i^{k-j} = \frac{1}{j!} \sum_{k=j}^{+\infty} \frac{c_k \cdot k!}{(k-j)!} \lambda_i^{k-j} = \frac{1}{j!} \left(\sum_{k=j}^{+\infty} c_k z^k \right)^{(j)} \Bigg|_{z=\lambda_i},$$

因幂级数收敛其导数也收敛，故$a_j \quad (j = 0, 1, \cdots, n_i - 1)$收敛. 因此，$f(J_i) = \sum\limits_{k=0}^{+\infty} c_k J_i^k$收敛，且

$$f(J_i) = \sum_{k=0}^{+\infty} c_k J_i^k = \begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \dfrac{f''(\lambda_i)}{2!} & \cdots & \dfrac{f^{(n_i-1)}(\lambda_i)}{(n_i-1)!} \\ & f(\lambda_i) & f'(\lambda_i) & \cdots & \dfrac{f^{(n_i-2)}(\lambda_i)}{(n_i-2)!} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda_i) \\ & & & & f(\lambda_i) \end{bmatrix},$$

$$i = 1, \cdots, t.$$

从而$f(A) = P\mathrm{diag}(f(J_1), f(J_2), \cdots, f(J_t))P^{-1}$收敛. □

推论 6.3.1 若存在方阵A的从属范数$\|\cdot\|$使得$\|A\|$落在$\sum\limits_{k=0}^{+\infty} c_k z^k$的收敛域内，则$\sum\limits_{k=0}^{+\infty} c_k A^k$收敛.

由定理6.3.2, 显然幂级数$\mathrm{e}^A, \sin A, \cos A$对任意的$A \in \mathbb{R}^{n\times n}$都是收敛的. 并且通过直接计算可得下面的性质:

性质 6.3.1 设$A \in \mathbb{F}^{n\times n}$, 则

(i) $\sin(-A) = -\sin A$, $\cos(-A) = \cos A$;

(ii) $\mathrm{e}^{\mathrm{i}A} = \cos A + \mathrm{i}\sin A$;

(iii) $\cos A = \frac{1}{2}(\mathrm{e}^{\mathrm{i}A} + \mathrm{e}^{-\mathrm{i}A})$, $\sin A = \frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}A} - \mathrm{e}^{-\mathrm{i}A})$.

性质 6.3.2 设$A, B \in \mathbb{R}^{n\times n}$, 若$AB = BA$, 则

(i) $\mathrm{e}^A\mathrm{e}^B = \mathrm{e}^B\mathrm{e}^A = \mathrm{e}^{A+B}$;

(ii) e^A一定可逆且$\left(\mathrm{e}^A\right)^{-1} = \mathrm{e}^{-A}$;

(iii) $\left(\mathrm{e}^A\right)^m = \mathrm{e}^{mA}$;

(iv) $\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B$,

$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B$,

$\cos 2A = \cos^2 A - \sin^2 A$.

例 6.3.1 设四阶矩阵A的特征值为π, $-\pi$, 0, 0, 求$\mathrm{e}^A, \sin A, \cos A$.

解 因为A的特征值为π, $-\pi$, 0, 0, 故A的特征多项式为$f(\lambda) = \lambda^4 - \lambda^2\pi^2$. 由$f(A) = 0$得$A^4 - \pi^2A^2 = 0$, 即$A^4 = \pi^2A^2$. 因此

$$
\begin{aligned}
A^5 = &\ A^4A = \pi^{(5-3)}A^3,\\
A^6 = &\ A^4A^2 = \pi^{(6-2)}A^2,\\
A^7 = &\ A^5A^2 = \pi^{(7-3)}A^3,\\
A^8 = &\ A^6A^2 = \pi^{(8-2)}A^2,\\
&\ \vdots
\end{aligned}
$$

从而

$$
\begin{aligned}
\mathrm{e}^A = &\ 1 + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \frac{1}{4!}A^4 + \frac{1}{5!}A^5 + \cdots + \frac{1}{n!}A^n + \cdots\\
= &\ 1 + A + \left(\sum_{k=0}^{\infty}\frac{\pi^{2k}}{(2k+2)!}\right)A^2 + \left(\sum_{k=0}^{\infty}\frac{\pi^{2k}}{(2k+3)!}\right)A^3,\\
\sin A = &\ A - \frac{1}{3!}A^3 + \frac{1}{5!}A^5 - \frac{1}{7!}A^7 + \cdots + \frac{(-1)^k}{(2k+1)!}A^{2k+1} + \cdots\\
= &\ A - \frac{1}{\pi^3}A^3\pi\\
&\ + \frac{1}{\pi^3}A^3\left(\pi - \frac{1}{3!}\pi^3 + \frac{1}{5!}\pi^5 - \cdots + \frac{(-1)^k}{(2k+1)!}\pi^{2k+1} + \cdots\right)\\
= &\ A - \frac{1}{\pi^2}A^3,\\
\cos A = &\ 1 - \frac{1}{2!}A^2 + \frac{1}{4!}A^4 - \frac{1}{6!}A^6 + \cdots + \frac{(-1)^k}{(2k)!}A^{2k} + \cdots\\
= &\ 1 + \frac{1}{\pi^2}A^2\left[-1 + \left(1 - \frac{1}{2!}\pi^2 + \frac{1}{4!}\pi^4 + \cdots + \frac{(-1)^k}{(2k)!}\pi^{2k} + \cdots\right)\right]\\
= &\ 1 - \frac{2}{\pi^2}A^2.
\end{aligned}
$$

□

习 题

1. 设函数矩阵

$$A(t)=\begin{bmatrix}\sin t & \cos t\\ \cos t & \sin t\end{bmatrix},$$

求$\lim\limits_{t\to 0}A(t)$.

2. 构造一个收敛的2阶可逆阵序列, 但它的极限矩阵不可逆.

3. 已知

$$A_k=\begin{bmatrix}\dfrac{1}{2^k} & \dfrac{\pi}{4^k}\\ 0 & \dfrac{1}{(k+1)(k+2)}\end{bmatrix}\quad(k=0,1,\cdots),$$

研究矩阵级数$\sum\limits_{k=0}^{+\infty}A_k$的敛散性.

4. 设$P\in\mathbb{F}^{m\times m}$和$Q\in\mathbb{F}^{n\times n}$可逆, 则矩阵级数$\sum\limits_{k=0}^{+\infty}A^{(k)}$(绝对)收敛的充要条件是$\sum\limits_{k=0}^{+\infty}\left(PA^{(k)}Q\right)$(绝对)收敛.

5. 若A为实反对称矩阵, 则e^{A}为正交阵.

6. 设A满足下面(i)~(iii)之一, 求e^{A^2}, $\sin(A)$, $\cos(A)$.
   (i) $A=A^2$; (ii) $A^2=I$; (iii) $A^2=0$.

7. 设$A=\begin{bmatrix}1&0&0&-1\\0&1&-1&0\\0&-1&1&0\\-1&0&0&1\end{bmatrix}$, 求$\mathrm{e}^{At}$.

8. 已知函数矩阵$\mathrm{e}^{At}=\begin{bmatrix}2\mathrm{e}^{2t}-\mathrm{e}^{t} & \mathrm{e}^{2t}-\mathrm{e}^{t} & \mathrm{e}^{t}-\mathrm{e}^{2t}\\ \mathrm{e}^{2t}-\mathrm{e}^{t} & 2\mathrm{e}^{2t}-\mathrm{e}^{t} & \mathrm{e}^{t}-\mathrm{e}^{2t}\\ 3\mathrm{e}^{2t}-3\mathrm{e}^{t} & 3\mathrm{e}^{2t}-3\mathrm{e}^{t} & 3\mathrm{e}^{t}-2\mathrm{e}^{2t}\end{bmatrix}$, 求矩阵$A$.

9. 已知

$$\sin(At)=\frac{1}{4}\begin{bmatrix}\sin 5t+3\sin t & 2\sin 5t-2\sin t & \sin 5t-\sin t\\ \sin 5t-\sin t & 2\sin 5t+2\sin t & \sin 5t-\sin t\\ \sin 5t-\sin t & 2\sin 5t-2\sin t & \sin 5t+3\sin t\end{bmatrix},$$

求A.

10. 若$A \in \mathbb{F}^{n\times n}$, 则 $\det(\mathrm{e}^{A}) = \mathrm{e}^{\mathrm{tr}A}$.
    (提示: 使用A的Jordan标准形.)

11. 若$A \in \mathbb{F}^{n\times n}$, 则$\|\mathrm{e}^{A}\| \leqslant \mathrm{e}^{\|A\|}$.

12. 设$\|\cdot\|$是$\mathbb{F}^{n\times n}$上的从属范数, 且$\|A\| < 1$, 则

    (i) $I_n + A$非奇异;

    (ii) $\|(I_n + A)^{-1}\| \leqslant (1 - \|A\|)^{-1}$.

13. 设A为n阶阵, 且A的所有特征值的实部均小于零, 则$\lim\limits_{t\to+\infty} e^{At} = 0$.

第7章　函数矩阵的微积分

本章将纯量函数的连续、导数和微积分等概念推广到函数矩阵. 主要介绍函数矩阵对单变量的导数、纯量函数对矩阵变量的导数、函数矩阵对矩阵变量的导数以及函数矩阵的微积分.

7.1　函数矩阵对单变量的导数

定义 7.1.1　若$m \times n$矩阵A的每个元素a_{ij}都是实变量t的函数, 称A为关于变量t的函数矩阵, 记作$[a_{ij}(t)]_{m\times n}$, 简记为$A(t)$.

定义 7.1.2　设函数矩阵$A(t) = [a_{ij}(t)]_{m\times n}$, 若对任意的$1 \leqslant i \leqslant m$和$1 \leqslant j \leqslant n$都有$a_{ij}(t)$在$t = t_0$处可导, 称$A(t)$在$t = t_0$处可导, 并称

$$\begin{bmatrix} a'_{11}(t_0) & a'_{12}(t_0) & \cdots & a'_{1n}(t_0) \\ a'_{21}(t_0) & a'_{22}(t_0) & \cdots & a'_{2n}(t_0) \\ \vdots & \vdots & \ddots & \vdots \\ a'_{m1}(t_0) & a'_{m2}(t_0) & \cdots & a'_{mn}(t_0) \end{bmatrix}$$

为$A(t)$在点t_0处的导数, 记作$\dfrac{\mathrm{d}A(t)}{\mathrm{d}t}\big|_{t=t_0}$或$A'(t_0)$.

定义 7.1.3　若函数矩阵$A(t)$的所有元素均在区间I上可导, 称$A(t)$在区间I上可导, 并称

$$\begin{bmatrix} a'_{11}(t) & a'_{12}(t) & \cdots & a'_{1n}(t) \\ a'_{21}(t) & a'_{22}(t) & \cdots & a'_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ a'_{m1}(t) & a'_{m2}(t) & \cdots & a'_{mn}(t) \end{bmatrix}, \ t \in I$$

为$A(t)$在区间I上的导数, 记作$\frac{\mathrm{d}A(t)}{\mathrm{d}t}, t \in I$或$A'(t), t \in I$.

例 7.1.1 设$A(t) = \begin{bmatrix} \mathrm{e}^t & t^2\mathrm{e}^t \\ t^3+1 & 0 \end{bmatrix}$, 则$A'(t) = \begin{bmatrix} \mathrm{e}^t & 2t\mathrm{e}^t + t^2\mathrm{e}^t \\ 3t^2 & 0 \end{bmatrix}$.

例 7.1.2 求函数矩阵e^{At}, $\cos(At)$, $\sin(At)$的导数.

解 由$\mathrm{e}^{At} = I + At + \frac{(At)^2}{2!} + \cdots + \frac{(At)^n}{n!} + \cdots$得

$$\begin{aligned} \frac{\mathrm{d}\mathrm{e}^{At}}{\mathrm{d}t} &= A + A^2t + \frac{A^3t^2}{2!} + \cdots + \frac{A^nt^{n-1}}{(n-1)!} + \cdots \\ &= A\left(I + At + \frac{(At)^2}{2!} + \cdots + \frac{(At)^n}{n!} + \cdots\right) \\ &= A\mathrm{e}^{At}. \end{aligned}$$

(注意: 上式的第一个等号不是对和式中的每一项求导再作和, 而是基于A的Jordan标准形和定义7.1.3得到.) 同理

$$\frac{\mathrm{d}\cos(At)}{\mathrm{d}t} = -A\sin(At), \quad \frac{\mathrm{d}\sin(At)}{\mathrm{d}t} = A\cos(At).$$

□

从上面的定义易见函数矩阵的导数有如下性质.

性质 7.1.1 设$A(t)$和$B(t)$为可导的函数矩阵, k为纯量, 则

(i) $[A(t) \pm B(t)]' = A'(t) \pm B'(t)$;

(ii) $[kA(t)]' = kA'(t)$;

(iii) $[A(t)B(t)]' = A'(t)B(t) + A(t)B'(t)$;

(iv) $[A^k(t)]' = \sum\limits_{i=0}^{k-1} A^i(t)A'(t)A^{k-1-i}(t)$.

性质 7.1.2 (复合函数矩阵求导公式) 设函数矩阵$A(t) = [a_{ij}(t)]_{m\times n}$, $t = g(x)$为纯量函数, 则

$$\frac{\mathrm{d}A(g(x))}{\mathrm{d}x} = \frac{\mathrm{d}A(g(x))}{\mathrm{d}g(x)} \cdot \frac{\mathrm{d}g(x)}{\mathrm{d}x}.$$

定义 7.1.4 (函数矩阵的高阶导数) 设函数矩阵$A(t)=[a_{ij}(t)]_{m\times n}$, 若对任意的$i,j$都有$a_{ij}(t)$在区间$I$上有$k$阶导数, 称

$$\begin{bmatrix} \dfrac{\mathrm{d}^k a_{11}(t)}{\mathrm{d}t^k} & \dfrac{\mathrm{d}^k a_{12}(t)}{\mathrm{d}t^k} & \cdots & \dfrac{\mathrm{d}^k a_{1n}(t)}{\mathrm{d}t^k} \\ \dfrac{\mathrm{d}^k a_{21}(t)}{\mathrm{d}t^k} & \dfrac{\mathrm{d}^k a_{22}(t)}{\mathrm{d}t^k} & \cdots & \dfrac{\mathrm{d}^k a_{2n}(t)}{\mathrm{d}t^k} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\mathrm{d}^k a_{m1}(t)}{\mathrm{d}t^k} & \dfrac{\mathrm{d}^k a_{m2}(t)}{\mathrm{d}t^k} & \cdots & \dfrac{\mathrm{d}^k a_{mn}(t)}{\mathrm{d}t^k} \end{bmatrix},\ t\in I$$

为$A(t)$的k阶导数, 记作$\dfrac{\mathrm{d}^k A(t)}{\mathrm{d}t^k}$.

从这个定义易见

$$\frac{\mathrm{d}^s A(t)}{\mathrm{d}t^s}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}^{s-1}A(t)}{\mathrm{d}t^{s-1}}\right),\ s=2,3,\cdots,k.$$

例 7.1.3 设$A(t)=\begin{bmatrix} t^2 & 0 \\ 1 & t \end{bmatrix}$, 求$\dfrac{\mathrm{d}^2 A(t)}{\mathrm{d}t^2}$.

解 由定义7.1.2及定义7.1.4得

$$\frac{\mathrm{d}A(t)}{\mathrm{d}t}=\begin{bmatrix} 2t & 0 \\ 0 & 1 \end{bmatrix},\quad \frac{\mathrm{d}^2A(t)}{\mathrm{d}t^2}=\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}.$$

□

例 7.1.4 设函数矩阵$A(t)=[a_{ij}(t)]_{n\times n}$, 证明: $\dfrac{\mathrm{d}\,(\mathrm{tr}A(t))}{\mathrm{d}t}=\mathrm{tr}\dfrac{\mathrm{d}A(t)}{\mathrm{d}t}$.

证明 $$\frac{\mathrm{d}\,(\mathrm{tr}A(t))}{\mathrm{d}t}=\frac{\mathrm{d}\left(\sum\limits_{i=1}^{n}a_{ii}(t)\right)}{\mathrm{d}t}=\sum_{i=1}^{n}\frac{\mathrm{d}a_{ii}(t)}{\mathrm{d}t}=\mathrm{tr}\frac{\mathrm{d}A(t)}{\mathrm{d}t}.$$ □

上面的例子表明求导运算和求迹运算可交换.

定义 7.1.5 对于函数矩阵$A(t)$, 若存在函数矩阵$B(t)$使得

$$A(t)B(t)=B(t)A(t)=I_n,$$

称函数矩阵$A(t)$可逆; 并称$B(t)$为$A(t)$的逆矩阵, 记为$A^{-1}(t)$.

例 7.1.5 设函数矩阵$A(t)$可逆, 且$A(t)$和$A^{-1}(t)$可导, 求$[A^{-1}(t)]'$.

解 由$A(t)A^{-1}(t)=I_n$及性质7.1.1得

$$A'(t)A^{-1}(t)+A(t)[A^{-1}(t)]'=0.$$

故$[A^{-1}(t)]'=-A^{-1}(t)A'(t)A^{-1}(t)$. □

7.2 纯量函数对矩阵变量的导数

定义 7.2.1 设$X=[x_{ij}]_{m\times n}\in\mathbb{R}^{m\times n}$, $f(X)$为$\mathbb{R}^{m\times n}$上的实函数, 若对任意的i,j都有偏导数$\dfrac{\partial f}{\partial x_{ij}}$存在, 称

$$\begin{bmatrix} \dfrac{\partial f}{\partial x_{11}} & \dfrac{\partial f}{\partial x_{12}} & \cdots & \dfrac{\partial f}{\partial x_{1n}} \\ \dfrac{\partial f}{\partial x_{21}} & \dfrac{\partial f}{\partial x_{22}} & \cdots & \dfrac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f}{\partial x_{m1}} & \dfrac{\partial f}{\partial x_{m2}} & \cdots & \dfrac{\partial f}{\partial x_{mn}} \end{bmatrix}$$

为f关于矩阵X的导数, 记作$\dfrac{\mathrm{d}f(X)}{\mathrm{d}X}$.

性质 7.2.1 设$f(x)$和$g(x)$是$\mathbb{R}^{m\times n}$上的纯量函数, $x\in\mathbb{R}^{m\times n}$, k是常数, 则

(i) $\dfrac{\mathrm{d}\,(f(x)\pm g(x))}{\mathrm{d}x}=\dfrac{\mathrm{d}f(x)}{\mathrm{d}x}\pm\dfrac{\mathrm{d}g(x)}{\mathrm{d}x}$;

(ii) $\dfrac{\mathrm{d}\,(kf(x))}{\mathrm{d}x}=k\dfrac{\mathrm{d}f(x)}{\mathrm{d}x}$;

(iii) $\dfrac{\mathrm{d}[f(x)g(x)]}{\mathrm{d}x}=\dfrac{\mathrm{d}f(x)}{\mathrm{d}x}g(x)+f(x)\dfrac{\mathrm{d}g(x)}{\mathrm{d}x}$.

定理 7.2.1 设$f(x)$为$\mathbb{R}^n$上的实纯量函数, $x(t)$是$n\times 1$实函数矩阵, 则

$$\frac{\mathrm{d}f(x(t))}{\mathrm{d}t}=\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^T\frac{\mathrm{d}f}{\mathrm{d}x}=\left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)^T\frac{\mathrm{d}x}{\mathrm{d}t}.$$

证明 设

$$x(t)=\begin{bmatrix}x_1(t) & x_2(t) & \cdots & x_n(t)\end{bmatrix}^T,$$

其中$x_1(t), x_2(t), \cdots, x_n(t)$为实纯量函数, 则

$$\begin{aligned}\frac{\mathrm{d}f(x(t))}{\mathrm{d}t} &= \frac{\partial f}{\partial x_1}\cdot\frac{\mathrm{d}x_1}{\mathrm{d}t}+\frac{\partial f}{\partial x_2}\cdot\frac{\mathrm{d}x_2}{\mathrm{d}t}+\cdots+\frac{\partial f}{\partial x_n}\cdot\frac{\mathrm{d}x_n}{\mathrm{d}t}\\ &= \begin{bmatrix}\frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n}\end{bmatrix}\begin{bmatrix}\frac{\mathrm{d}x_1}{\mathrm{d}t}\\ \frac{\mathrm{d}x_2}{\mathrm{d}t}\\ \vdots\\ \frac{\mathrm{d}x_n}{\mathrm{d}t}\end{bmatrix}\\ &= \left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)^T\cdot\frac{\mathrm{d}x}{\mathrm{d}t}\\ &= \left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^T\cdot\frac{\mathrm{d}f}{\mathrm{d}x}.\end{aligned}$$

□

例 7.2.1 设$A=[a_{ij}]_{n\times n}\in\mathbb{R}^{n\times n}$, $x\in\mathbb{R}^n$, 求$\frac{\mathrm{d}(x^TAx)}{\mathrm{d}x}$, $\frac{\mathrm{d}(x^TAx)}{\mathrm{d}A}$. 特别地, 若$A=I_n$, 则$\frac{\mathrm{d}(x^Tx)}{\mathrm{d}x}=2x$; 若$A=A^T$, 则$\frac{\mathrm{d}(x^TAx)}{\mathrm{d}x}=2Ax$.

解 令$x=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\end{bmatrix}^T$, 由$x^TA\frac{\partial x}{\partial x_i}$为纯量函数及

$$\left(x^TA\frac{\partial x}{\partial x_i}\right)^T=\left(\frac{\partial x}{\partial x_i}\right)^TA^Tx=\frac{\partial x^T}{\partial x_i}A^Tx$$

推出$x^TA\frac{\partial x}{\partial x_i}=\frac{\partial x^T}{\partial x_i}A^Tx$. 所以

$$\frac{\partial(x^TAx)}{\partial x_i}=\frac{\partial x^T}{\partial x_i}(A+A^T)x,\ i=1,2,\cdots,n,$$

故

$$\frac{\mathrm{d}\left(x^TAx\right)}{\mathrm{d}x}=\begin{bmatrix}\frac{\partial\left(x^TAx\right)}{\partial x_1}\\ \frac{\partial\left(x^TAx\right)}{\partial x_2}\\ \vdots\\ \frac{\partial\left(x^TAx\right)}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\frac{\partial x^T}{\partial x_1}\\ \frac{\partial x^T}{\partial x_2}\\ \vdots\\ \frac{\partial x^T}{\partial x_n}\end{bmatrix}\left(A+A^T\right)x=\left(A+A^T\right)x.$$

由

$$\frac{\partial(x^TAx)}{\partial a_{ij}}=x^TE_{ij}x=x_ix_j,\quad i,j=1,2,\cdots,n,$$

得

$$\frac{\mathrm{d}(x^TAx)}{\mathrm{d}A}=\begin{bmatrix}\frac{\partial\left(x^TAx\right)}{\partial a_{11}} & \frac{\partial\left(x^TAx\right)}{\partial a_{12}} & \cdots & \frac{\partial\left(x^TAx\right)}{\partial a_{1n}}\\ \frac{\partial\left(x^TAx\right)}{\partial a_{21}} & \frac{\partial\left(x^TAx\right)}{\partial a_{22}} & \cdots & \frac{\partial\left(x^TAx\right)}{\partial a_{2n}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial\left(x^TAx\right)}{\partial a_{n1}} & \frac{\partial\left(x^TAx\right)}{\partial a_{n2}} & \cdots & \frac{\partial\left(x^TAx\right)}{\partial a_{nn}}\end{bmatrix}=xx^T.$$

□

例 7.2.2 设$A\in\mathbb{R}^{n\times n}$, 证明: $\det\left(\frac{\mathrm{d}(\det A)}{\mathrm{d}A}\right)=(\det A)^{n-1}$.

证明 设$A=[a_{ij}]$, 易见$\frac{\partial(\det A)}{\partial a_{ij}}=A_{ji}$, $i,j=1,2,\cdots,n$, 其中A_{ij}为矩阵A的(i,j)位置的代数余子式, 于是

$$\frac{\mathrm{d}(\det A)}{\mathrm{d}A}=\begin{bmatrix}\frac{\partial(\det A)}{\partial a_{11}} & \frac{\partial(\det A)}{\partial a_{12}} & \cdots & \frac{\partial(\det A)}{\partial a_{1n}}\\ \frac{\partial(\det A)}{\partial a_{21}} & \frac{\partial(\det A)}{\partial a_{22}} & \cdots & \frac{\partial(\det A)}{\partial a_{2n}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial(\det A)}{\partial a_{n1}} & \frac{\partial(\det A)}{\partial a_{n2}} & \cdots & \frac{\partial(\det A)}{\partial a_{nn}}\end{bmatrix}=(\mathrm{adj}A)^T,$$

从而

$$\det\left(\frac{\mathrm{d}(\det A)}{\mathrm{d}A}\right)=\det\left((\mathrm{adj}A)^T\right)=\det(\mathrm{adj}A)=(\det A)^{n-1}.$$

□

例 7.2.3 设$X=[x_{ij}]\in\mathbb{R}^{n\times n}$, 求$\frac{\mathrm{d}(\mathrm{tr}X)}{\mathrm{d}X}$, $\frac{\mathrm{d}\left(\mathrm{tr}(X^TX)\right)}{\mathrm{d}X}$, $\frac{\mathrm{d}(\mathrm{tr}X^2)}{\mathrm{d}X}$.

解 易见

$$\frac{\mathrm{d}(\mathrm{tr}X)}{\mathrm{d}X}=\begin{bmatrix}\frac{\partial(\mathrm{tr}X)}{\partial x_{11}} & \frac{\partial(\mathrm{tr}X)}{\partial x_{12}} & \cdots & \frac{\partial(\mathrm{tr}X)}{\partial x_{1n}}\\ \frac{\partial(\mathrm{tr}X)}{\partial x_{21}} & \frac{\partial(\mathrm{tr}X)}{\partial x_{22}} & \cdots & \frac{\partial(\mathrm{tr}X)}{\partial x_{2n}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial(\mathrm{tr}X)}{\partial x_{n1}} & \frac{\partial(\mathrm{tr}X)}{\partial x_{n2}} & \cdots & \frac{\partial(\mathrm{tr}X)}{\partial x_{nn}}\end{bmatrix}=I_n.$$

因为$\mathrm{tr}(X^TX)=\sum\limits_{i,j=1}^{n}x_{ij}^2$, 所以

$$\frac{\mathrm{d}\left(\mathrm{tr}(X^TX)\right)}{\mathrm{d}X}=\frac{\mathrm{d}\left(\sum\limits_{i,j=1}^{n}x_{ij}^2\right)}{\mathrm{d}X}=\begin{bmatrix}2x_{11} & 2x_{12} & \cdots & 2x_{1n}\\ 2x_{21} & 2x_{22} & \cdots & 2x_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ 2x_{n1} & 2x_{n2} & \cdots & 2x_{nn}\end{bmatrix}=2X.$$

由求导运算和求逆运算可交换得

$$\begin{aligned}\frac{\partial(\mathrm{tr}X^2)}{\partial x_{ij}} &= \mathrm{tr}\frac{\partial X^2}{\partial x_{ij}}=\mathrm{tr}\left(\frac{\partial X}{\partial x_{ij}}X+X\frac{\partial X}{\partial x_{ij}}\right)\\ &= \mathrm{tr}\left(E_{ij}X+XE_{ij}\right)=2x_{ji},\quad \forall i,j,\end{aligned}$$

所以

$$\frac{\mathrm{d}(\mathrm{tr}X^2)}{\mathrm{d}X}=\begin{bmatrix}\frac{\partial(\mathrm{tr}X^2)}{\partial x_{11}} & \frac{\partial(\mathrm{tr}X^2)}{\partial x_{12}} & \cdots & \frac{\partial(\mathrm{tr}X^2)}{\partial x_{1n}}\\ \frac{\partial(\mathrm{tr}X^2)}{\partial x_{21}} & \frac{\partial(\mathrm{tr}X^2)}{\partial x_{22}} & \cdots & \frac{\partial(\mathrm{tr}X^2)}{\partial x_{2n}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial(\mathrm{tr}X^2)}{\partial x_{n1}} & \frac{\partial(\mathrm{tr}X^2)}{\partial x_{n2}} & \cdots & \frac{\partial(\mathrm{tr}X^2)}{\partial x_{nn}}\end{bmatrix}=2X^T.$$

□

7.3 函数矩阵对矩阵变量的导数

定义 7.3.1 设$A(t)=[a_{ij}(t)]_{m\times n}$为函数矩阵, $a_{ij}(t)$为实纯量函数, $B=[b_{ij}]_{p\times q}\in\mathbb{R}^{p\times q}$, 称

$$\begin{bmatrix}\frac{\partial A(t)}{\partial b_{11}} & \frac{\partial A(t)}{\partial b_{12}} & \cdots & \frac{\partial A(t)}{\partial b_{1q}}\\ \frac{\partial A(t)}{\partial b_{21}} & \frac{\partial A(t)}{\partial b_{22}} & \cdots & \frac{\partial A(t)}{\partial b_{2q}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial A(t)}{\partial b_{p1}} & \frac{\partial A(t)}{\partial b_{p2}} & \cdots & \frac{\partial A(t)}{\partial b_{pq}}\end{bmatrix}$$

为$A(t)$对B的导数, 记作$\frac{\mathrm{d}A(t)}{\mathrm{d}B}$.

例 7.3.1 设$x \in \mathbb{R}^n$, 求$\dfrac{\mathrm{d}x^T}{\mathrm{d}x}$.

解 设$x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$, 则

$$\frac{\mathrm{d}x^T}{\mathrm{d}x} = \begin{bmatrix} \dfrac{\partial x^T}{\partial x_1} \\ \dfrac{\partial x^T}{\partial x_2} \\ \vdots \\ \dfrac{\partial x^T}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial x_1}{\partial x_1} & \dfrac{\partial x_2}{\partial x_1} & \cdots & \dfrac{\partial x_n}{\partial x_1} \\ \dfrac{\partial x_1}{\partial x_2} & \dfrac{\partial x_2}{\partial x_2} & \cdots & \dfrac{\partial x_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial x_1}{\partial x_n} & \dfrac{\partial x_2}{\partial x_n} & \cdots & \dfrac{\partial x_n}{\partial x_n} \end{bmatrix} = I_n.$$

□

例 7.3.2 设

$$x = [x_1\ x_2\ \cdots\ x_n],\ f(x) = \begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_m(x) \end{bmatrix}^T,$$

$f_1(x), f_2(x), \cdots, f_m(x)$均为纯量函数, 证明:

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \begin{bmatrix} \dfrac{\partial f_1(x)}{\partial x_1} & \dfrac{\partial f_1(x)}{\partial x_2} & \cdots & \dfrac{\partial f_1(x)}{\partial x_n} \\ \dfrac{\partial f_2(x)}{\partial x_1} & \dfrac{\partial f_2(x)}{\partial x_2} & \cdots & \dfrac{\partial f_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m(x)}{\partial x_1} & \dfrac{\partial f_m(x)}{\partial x_2} & \cdots & \dfrac{\partial f_m(x)}{\partial x_n} \end{bmatrix}.$$

证明

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \left[\frac{\partial f}{\partial x_1}\ \frac{\partial f}{\partial x_2}\ \cdots\ \frac{\partial f}{\partial x_n}\right] = \begin{bmatrix} \dfrac{\partial f_1(x)}{\partial x_1} & \dfrac{\partial f_1(x)}{\partial x_2} & \cdots & \dfrac{\partial f_1(x)}{\partial x_n} \\ \dfrac{\partial f_2(x)}{\partial x_1} & \dfrac{\partial f_2(x)}{\partial x_2} & \cdots & \dfrac{\partial f_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m(x)}{\partial x_1} & \dfrac{\partial f_m(x)}{\partial x_2} & \cdots & \dfrac{\partial f_m(x)}{\partial x_n} \end{bmatrix}.$$

□

例 7.3.3 设$A \in \mathbb{R}^{n\times n}$, $y \in \mathbb{R}^n$, 求$\dfrac{\mathrm{d}(y^TA^TAy)}{\mathrm{d}A}$.

解 设$y = [y_1\ y_2\ \cdots\ y_n]^T, A = [a_{ij}]$, 则

$$\begin{aligned} \frac{\partial\left(y^TA^TAy\right)}{\partial a_{ij}} &= y^T\frac{\partial A^T}{\partial a_{ij}}Ay + y^TA^T\frac{\partial A}{\partial a_{ij}}y \\ &= y^TE_{ji}Ay + y^TA^TE_{ij}y \\ &= 2y^TE_{ji}Ay \\ &= 2y_ja_iy = 2a_iyy_j,\quad \forall i, j, \end{aligned}$$

其中a_i为A的第i行, 故

$$\frac{\mathrm{d}(y^TA^TAy)}{\mathrm{d}A}=\begin{bmatrix}2a_1yy_1 & 2a_1yy_2 & \cdots & 2a_1yy_n\\ 2a_2yy_1 & 2a_2yy_2 & \cdots & 2a_2yy_n\\ \vdots & \vdots & \ddots & \vdots\\ 2a_nyy_1 & 2a_nyy_2 & \cdots & 2a_nyy_n\end{bmatrix}=2Ayy^T.$$

□

例 7.3.4 设$X(u)$为$n\times 1$阶实函数矩阵, $u\in\mathbb{R}^m$, $f(X)$为$\mathbb{R}^n$上的实纯量函数, 证明: $\frac{\mathrm{d}f(X)}{\mathrm{d}u}=\frac{\mathrm{d}X^T(u)}{\mathrm{d}u}\cdot\frac{\mathrm{d}f(X)}{\mathrm{d}X}$.

证明 设

$$X(u)=\begin{bmatrix}x_1(u) & x_2(u) & \cdots & x_n(u)\end{bmatrix}^T,$$

$$u=\begin{bmatrix}u_1 & u_2 & \cdots & u_m\end{bmatrix}^T,$$

则

$$\begin{aligned}\frac{\mathrm{d}f(X)}{\mathrm{d}u} &= \begin{bmatrix}\frac{\partial f(X)}{\partial u_1} & \frac{\partial f(X)}{\partial u_2} & \cdots & \frac{\partial f(X)}{\partial u_m}\end{bmatrix}^T\\ &= \begin{bmatrix}\frac{\partial f(X)}{\partial x_1}\cdot\frac{\partial x_1}{\partial u_1}+\frac{\partial f(X)}{\partial x_2}\cdot\frac{\partial x_2}{\partial u_1}+\cdots+\frac{\partial f(X)}{\partial x_n}\cdot\frac{\partial x_n}{\partial u_1}\\ \frac{\partial f(X)}{\partial x_1}\cdot\frac{\partial x_1}{v\partial u_2}+\frac{\partial f(X)}{\partial x_2}\cdot\frac{\partial x_2}{\partial u_2}+\cdots+\frac{\partial f(X)}{\partial x_n}\cdot\frac{\partial x_n}{\partial u_2}\\ \vdots\\ \frac{\partial f(X)}{\partial x_1}\cdot\frac{\partial x_1}{\partial u_m}+\frac{\partial f(X)}{\partial x_2}\cdot\frac{\partial x_2}{\partial u_m}+\cdots+\frac{\partial f(X)}{\partial x_n}\cdot\frac{\partial x_n}{\partial u_m}\end{bmatrix}\\ &= \begin{bmatrix}\frac{\mathrm{d}x_1}{\mathrm{d}u} & \frac{\mathrm{d}x_2}{\mathrm{d}u} & \cdots & \frac{\mathrm{d}x_n}{\mathrm{d}u}\end{bmatrix}\frac{\mathrm{d}f(X)}{\mathrm{d}X}\\ &= \frac{\mathrm{d}X^T(u)}{\mathrm{d}u}\cdot\frac{\mathrm{d}f(X)}{\mathrm{d}X}.\end{aligned}$$

□

7.4 函数矩阵的微积分

定义 7.4.1 若函数矩阵$A(t)$的所有元素在$t=t_0$处连续(右连续, 左连续), 称$A(t)$在$t=t_0$处连续(右连续, 左连续).

定义 7.4.2 若函数矩阵$A(t)$在(a,b)内每一点都连续(右连续, 左连续), 称$A(t)$在(a,b)内连续(右连续, 左连续).

定义 7.4.3 若函数矩阵$A(t)$在(a,b)内连续, 并且在a点右连续, 在b点左连续, 称$A(t)$在$[a,b]$上连续.

可类似上面定义函数矩阵的微分和积分, 即:

定义 7.4.4 设函数矩阵$A(t)$中所有元素$a_{ij}(t)$都在$t=t_0$处(或某区间内)可微, 称矩阵$A(t)$在$t=t_0$处(或某区间内)可微.

定义 7.4.5 设$A(t)=[a_{ij}(t)]_{m\times n}$, 若对任意的$i,j$, $a_{ij}(t)$在区间$[a,b]$上可积, 称$A(t)$在区间$[a,b]$上可积; 并称

$$\begin{bmatrix} \int_a^b a_{11}(t)\mathrm{d}t & \int_a^b a_{12}(t)\mathrm{d}t & \cdots & \int_a^b a_{1n}(t)\mathrm{d}t \\ \int_a^b a_{21}(t)\mathrm{d}t & \int_a^b a_{22}(t)\mathrm{d}t & \cdots & \int_a^b a_{2n}(t)\mathrm{d}t \\ \vdots & \vdots & \ddots & \vdots \\ \int_a^b a_{m1}(t)\mathrm{d}t & \int_a^b a_{m2}(t)\mathrm{d}t & \cdots & \int_a^b a_{mn}(t)\mathrm{d}t \end{bmatrix}$$

为$A(t)$在区间$[a,b]$上的积分, 记作$\int_a^b A(t)\mathrm{d}t$.

性质 7.4.1 若$A(x)$和$B(x)$是$[t_0,t]$上有适当阶数的可积函数矩阵, k为常数矩阵, 则

(i) $\int_{t_0}^t [A(x)\pm B(x)]\mathrm{d}x=\int_{t_0}^t A(x)\mathrm{d}x\pm\int_{t_0}^t B(x)\mathrm{d}x$;

(ii) $\int_{t_0}^t kA(x)\mathrm{d}x=k\int_{t_0}^t A(x)\mathrm{d}x$;

(iii) 当$A(x)$在$[t_0,t]$上连续时有$\left(\int_{t_0}^t A(x)\mathrm{d}x\right)'=A(t)$;

(iv) 当$A'(x)$在$[t_0,t]$上连续时有$\int_{t_0}^t A'(x)\mathrm{d}x=A(t)-A(t_0)$.

习 题

1. 设函数矩阵$A(t) = \begin{bmatrix} \sin t & \cos t & t \\ \frac{\sin t}{t} & e^t & t^2 \\ 1 & 0 & t^2 \end{bmatrix}$, $t \neq 0$, 计算$A'(t)$, $A''(t)$, $\det[A(t)]''$, $\det[A''(t)]$.

2. 设$A(t)$是n阶函数矩阵. 举例说明, 对正整数$m > 1$, 关系式

$$\frac{d}{dt}(A(t))^m = m(A(t))^{m-1}\frac{d}{dt}A(t)$$

一般不成立.

3. 设$A(t) = \prod_{i=1}^{n} A_i(t)$, 其中$A_i(t)$为$n \times n$阶的函数矩阵, 证明:

$$\frac{dA(t)}{dt} = \sum_{i=1}^{n}\left(A_1(t)\ \cdots\ A_{i-1}(t)\ \frac{dA_i(t)}{dt}\ A_{i+1}(t)\ \cdots\ A_n(t)\right).$$

4. 设$x = [x_1\ x_2\ \cdots\ x_n]^T$, $f(x) = x^T x$, 求$\frac{df(x)}{dx}$, $\frac{df(x)}{dx^T}$.

5. 设$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix}$, $Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$, 求$\frac{d(AY)}{dA}$, $\frac{d(AY)^T}{dA}$.

6. 设$X \in \mathbb{R}^{n\times m}$, $A \in \mathbb{R}^{m\times n}$, 求证:

$$\frac{d}{dX}(\operatorname{tr}(AX)) = \frac{d}{dX}\left(\operatorname{tr}(X^T A^T)\right) = A^T.$$

7. 若$A(s)$为多项式矩阵, 则$\frac{d(\det A(s))}{ds} = \operatorname{tr}\left(\operatorname{adj}A(s)\frac{dA(s)}{ds}\right)$.

(提示: 使用行列式的定义.)

8. 设函数矩阵$A(t) = \begin{bmatrix} \sin t & \cos t \\ \cos t & \sin t \end{bmatrix}$, 求$\int_0^t A(s)ds$, $\frac{d}{dt}\left(\int_0^{t^2} A(s)ds\right)$.

9. 设函数矩阵

$$A(t) = \begin{bmatrix} e^{2t} & te^t & t^2 \\ e^{-t} & 2e^{2t} & 0 \\ t\sin t & 0 & 0 \end{bmatrix},$$

求$\int_0^1 A(t)dt$, $\frac{d}{dt}\left(\int_0^{t^2} A(t)\right)$.

第8章 矩阵的特征值和奇异值不等式

本章介绍Courant-Fischer定理及其推广形式, 并由此给出关于矩阵特征值和奇异值的一些重要不等式. 8.3~8.5摘自参考文献[14].

8.1 Courant-Fischer定理及其应用

8.1.1 Courant-Fischer定理

借助于向量范数, 显然定理2.4.4的一个等价表述为:

定理 8.1.1 设$A^* = A \in \mathbb{C}^{n\times n}$, 则

$$\max_{\substack{y\in\mathbb{C}^n\\ \|y\|_2=1}} y^*Ay = \lambda_{\max}(A), \quad \min_{\substack{y\in\mathbb{C}^n\\ \|y\|_2=1}} y^*Ay = \lambda_{\min}(A).$$

推论 8.1.1 设$A^* = A = [a_{ij}] \in \mathbb{C}^{n\times n}$, 则

(i) $\lambda_{\min}(A) \leqslant a_{ii} \leqslant \lambda_{\max}(A)$, $i = 1, 2, \cdots, n$;

(ii) 当$A > 0$时, A的元素中模最大者一定落在A的对角线上.

证明 (i) 用e_i表示单位阵的第i列, $i = 1, 2, \cdots, n$, 则

$$\|e_i\|_2 = 1,\ (e_i)^T Ae_i = a_{ii},\ i = 1, 2, \cdots, n.$$

由定理8.1.1得(i)成立.

(ii) 对于任意互异的$1 \leqslant i, j \leqslant n$, 由$A > 0$得

$$a_{ii} > 0,\ a_{jj} > 0,\ \begin{bmatrix} a_{ii} & a_{ij} \\ \overline{a_{ij}} & a_{jj} \end{bmatrix} > 0,$$

于是

$$a_{ii}a_{jj} > a_{ij}\overline{a_{ij}} = |a_{ij}|^2,$$

从而

$$|a_{ij}| < \sqrt{a_{ii}a_{jj}} \leqslant \max\{a_{ii}, a_{jj}\}.$$

故A的元素中模最大者一定落在A的对角线上. □

定理 8.1.2 设$A^* = A \in \mathbb{C}^{n\times n}$, $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$是$A$的特征值, q_i是A的属于特征值λ_i的特征向量, $i = 1, 2, \cdots, n$, 并且向量组q_1, q_2, $\cdots$, q_n标准正交, 则

(i) $$\max_{\substack{x\perp \mathrm{span}\{q_1,\cdots,q_{i-1}\}\\ \|x\|_2=1}} x^*Ax = \lambda_i,\ i = 2, \cdots, n;$$

(ii) $$\min_{\substack{x\perp \mathrm{span}\{q_{i+1},\cdots,q_n\}\\ \|x\|_2=1}} x^*Ax = \lambda_i,\ i = 1, 2, \cdots, n-1.$$

证明 (i) 既然q_1, q_2, $\cdots$, q_n是$\mathbb{C}^n$的标准正交基, $\mathbb{C}^n$中的任意向量x均可表示为

$$x = x_1q_1 + x_2q_2 + \cdots + x_nq_n,\ x_i \in \mathbb{C},\ i = 1, 2, \cdots, n.$$

故

$$\begin{aligned}
&\max_{\substack{x\perp \mathrm{span}\{q_1,\cdots,q_{i-1}\}\\ \|x\|_2=1}} x^*Ax \\
=& \max_{\substack{x\perp \mathrm{span}\{q_1,\cdots,q_{i-1}\}\\ \|x\|_2=1}} (x_1q_1 + \cdots + x_nq_n)^*A(x_1q_1 + \cdots + x_nq_n) \\
=& \max_{\|\tilde{x}\|_2=1} (x_iq_i + \cdots + x_nq_n)^*A(x_iq_i + \cdots + x_nq_n) \\
=& \max_{\|\tilde{x}\|_2=1} \lambda_i|x_i|^2 + \cdots + \lambda_n|x_n|^2 \\
\leqslant& \lambda_i,
\end{aligned}$$

其中$\tilde{x} = \begin{bmatrix} x_i & \cdots & x_n \end{bmatrix}^T$. 取$x = q_i$, 则$x^*Ax = q_i^*Aq_i = \lambda_i q_i^* q_i = \lambda_i$, 故

$$\max_{\substack{x\perp \mathrm{span}\{q_1,\cdots,q_{i-1}\}\\ \|x\|_2=1}} x^*Ax = \lambda_i,\ i = 2, \cdots, n-1.$$

(ii) 类似于(i). □

定理 8.1.3 设$A^* = A \in \mathbb{C}^{n\times n}$, $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$是$A$的特征值, 则

(i) $$\min_{B\in\mathbb{C}_{i-1}^{n\times(i-1)}} \max_{\substack{B^*x=0\\ \|x\|_2=1}} x^*Ax = \lambda_i,\ i = 2, \cdots, n;$$

(ii) $$\max_{C\in\mathbb{C}_{n-i}^{n\times(n-i)}} \min_{\substack{C^*x=0\\ \|x\|_2=1}} x^*Ax = \lambda_i,\ i = 1, 2, \cdots, n-1.$$

证明 (i) 由$A=A^*$及正交对角化原理知

$$A=\begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}\operatorname{diag}(\lambda_1,\cdots,\lambda_n)\begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}^*, \tag{8.1.1}$$

其中$\begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}$是酉矩阵.

对于$i=2,\cdots,n$, 令$B_1=[q_1 \ \cdots \ q_{i-1}]$, 则

$$\min_{B\in\mathbb{C}_{i-1}^{n\times(i-1)}} \max_{\substack{B^*x=0\\ \|x\|_2=1}} x^*Ax \leqslant \max_{\substack{B_1^*x=0\\ \|x\|_2=1}} x^*Ax = \max_{\substack{[q_1 \ \cdots \ q_{i-1}]^*x=0\\ \|x\|_2=1}} x^*Ax .$$

由定理8.1.2(i)知

$$\min_{B\in\mathbb{C}_{i-1}^{n\times(i-1)}} \max_{\substack{B^*x=0\\ \|x\|_2=1}} x^*Ax \leqslant \lambda_i.$$

任取$B_0\in\mathbb{C}_{i-1}^{n\times(i-1)}$, 则$\mathcal{N}(B_0^*)\cap\operatorname{span}\{q_1,\cdots,q_i\}\neq\varnothing$, 于是存在非零的$z\in\operatorname{span}\{q_1,\cdots,q_i\}$使得$\|z\|_2=1$且$B_0^*z=0$, 故

$$\begin{aligned}\max_{\substack{B_0^*x=0\\ \|x\|_2=1}} x^*Ax \geqslant{}& z^*Az\\ \geqslant{}& \min_{\substack{w\in\operatorname{span}\{q_1,\cdots,q_i\}\\ \|w\|_2=1}} w^*Aw\\ ={}& \min_{\substack{v\in\mathbb{C}^i\\ \|v\|_2=1}} v^*\begin{bmatrix} q_1 & \cdots & q_i \end{bmatrix}^*A\begin{bmatrix} q_1 & \cdots & q_i \end{bmatrix}v.\end{aligned}$$

使用式(8.1.1)和定理8.1.1推出

$$\max_{\substack{B_0^*x=0\\ \|x\|_2=1}} x^*Ax \geqslant \min_{\substack{v\in\mathbb{C}^i\\ \|v\|_2=1}} v^*\operatorname{diag}(\lambda_1,\cdots,\lambda_i)v=\lambda_i.$$

由B_0的任意性知 $\displaystyle\min_{B\in\mathbb{C}_{i-1}^{n\times(i-1)}} \max_{\substack{B^*x=0\\ \|x\|_2=1}} x^*Ax \geqslant \lambda_i.$

总之

$$\min_{B\in\mathbb{C}_{i-1}^{n\times(i-1)}} \max_{\substack{B^*x=0\\ \|x\|_2=1}} x^*Ax = \lambda_i,\ i=2,\cdots,n.$$

(ii) 类似于(i). □

定理8.1.1和定理8.1.3可以被统一表示为:

定理 8.1.4 (Courant-Fischer**定理**) 设$A^*=A\in\mathbb{C}^{n\times n}$, $\lambda_1\geqslant\cdots\geqslant\lambda_n$是$A$的特征值, 则

(i) $\displaystyle\min_{B\in\mathbb{C}_{n-i+1}^{n\times(n-i+1)}} \max_{\substack{x\in\mathcal{R}(B)\\ \|x\|_2=1}} x^*Ax=\lambda_i,\ i=1,2,\cdots,n;$

(ii) $\displaystyle\max_{C\in\mathbb{C}_i^{n\times i}} \min_{\substack{x\in\mathcal{R}(C)\\ \|x\|_2=1}} x^*Ax=\lambda_i,\ i=1,2,\cdots,n.$

证明 证明留给读者. □

8.1.2 Sturm分离原理

下面应用Courant-Fischer定理(即定理8.1.4)证明Sturm分离原理.

定理 8.1.5 (Sturm分离原理) 设$A^* = A \in \mathbb{C}^{n\times n}$, A_t是A的$t(t < n)$阶顺序主子阵, $\lambda_1(A_t) \geqslant \lambda_2(A_t) \geqslant \cdots \geqslant \lambda_t(A_t)$是$A_t$的特征值, 则

$$\lambda_i(A_{t+1}) \geqslant \lambda_i(A_t) \geqslant \lambda_{i+1}(A_{t+1}), \quad i = 1, 2, \cdots, t.$$

证明 由Courant-Fischer定理, 存在$B_1 \in \mathbb{C}_{t-i+1}^{t\times(t-i+1)}$使得

$$\begin{aligned}
\lambda_i(A_t) = & \max_{\substack{x\in\mathcal{R}(B_1)\\ \|x\|_2=1}} x^* A_t x \\
= & \max_{\substack{x\in\mathcal{R}(B_1)\\ \|x\|_2=1}} [x^*\ 0] A_{t+1} \begin{bmatrix} x \\ 0 \end{bmatrix} \\
= & \max_{\substack{y\in\mathcal{R}(B_2)\\ \|y\|_2=1}} y^* A_{t+1} y \\
\geqslant & \min_{B\in\mathbb{C}_{t-i+1}^{(t+1)\times(t-i+1)}} \max_{\substack{y\in\mathcal{R}(B)\\ \|y\|_2=1}} y^* A_{t+1} y \\
= & \lambda_{i+1}(A_{t+1}),\ i = 1, \cdots, t,
\end{aligned}$$

其中$B_2 = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}$.

由Courant-Fischer定理, 存在$C_1 \in \mathbb{C}_i^{t\times i}$使得

$$\begin{aligned}
\lambda_i(A_t) = & \min_{\substack{x\in\mathcal{R}(C_1)\\ \|x\|_2=1}} x^* A_t x \\
= & \min_{\substack{x\in\mathcal{R}(C_1)\\ \|x\|_2=1}} [x^*\ 0] A_{t+1} \begin{bmatrix} x \\ 0 \end{bmatrix} \\
= & \min_{\substack{y\in\mathcal{R}(C_2)\\ \|y\|_2=1}} y^* A_{t+1} y \\
\leqslant & \max_{C\in\mathbb{C}_i^{(t+1)\times i}} \min_{\substack{y\in\mathcal{R}(C)\\ \|y\|_2=1}} y^* A_{t+1} y \\
= & \lambda_i(A_{t+1}),\ i = 1, \cdots, t,
\end{aligned}$$

其中$C_2 = \begin{bmatrix} C_1 \\ 0 \end{bmatrix}$. □

推论 8.1.2 设$A^* = A \in \mathbb{C}^{n\times n}$, A_k是A的k阶顺序主子阵,

$$\lambda_1(A_k) \geqslant \lambda_2(A_k) \geqslant \cdots \geqslant \lambda_k(A_k)$$

是A_k的特征值, 则

$$\lambda_i(A) \geqslant \lambda_i(A_k) \geqslant \lambda_{i+n-k}(A),\ i = 1, 2, \cdots, k.$$

证明 由Sturm分离原理知

$$\lambda_i(A) \geqslant \lambda_i(A_{n-1}) \geqslant \cdots \geqslant \lambda_i(A_k);$$

$$\lambda_i(A_k) \geqslant \lambda_{i+1}(A_{k+1}) \geqslant \cdots \geqslant \lambda_{i+n-k}(A),$$

故$\lambda_i(A) \geqslant \lambda_i(A_k) \geqslant \lambda_{i+n-k}(A), \quad i = 1, 2, \cdots, k.$ □

8.1.3 Weyl型不等式

定理 8.1.6 设A, $B \in \mathbb{C}^{n\times n}$满足$A^* = A$且$B^* = B$, 正整数$i, j$满足$i \leqslant n$且$j \leqslant n$, $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$是$A$的特征值, $\lambda_1(B) \geqslant \cdots \geqslant \lambda_n(B)$是$B$的特征值, $\lambda_1(A+B) \geqslant \cdots \geqslant \lambda_n(A+B)$是$A+B$的特征值, 则

(i) 当$i + j \geqslant n + 1$时, 有$\lambda_i(A) + \lambda_j(B) \leqslant \lambda_{i+j-n}(A+B)$;

(ii) 当$i + j \leqslant n + 1$时, 有$\lambda_i(A) + \lambda_j(B) \geqslant \lambda_{i+j-1}(A+B)$.

证明 (i) 由定理8.1.4知存在$C \in \mathbb{C}_i^{n\times i}$及$D \in \mathbb{C}_j^{n\times j}$使得

$$\lambda_i(A) = \min_{\substack{x\in\mathcal{R}(C)\\ \|x\|_2=1}} x^*Ax, \quad \lambda_j(B) = \min_{\substack{y\in\mathcal{R}(D)\\ \|y\|_2=1}} y^*By.$$

令$V = \mathcal{R}(C) \cap \mathcal{R}(D)$, 则

$$\dim V = \dim\mathcal{R}(C) + \dim\mathcal{R}(D) - \dim(\mathcal{R}(C) + \mathcal{R}(D)) \geqslant i + j - n.$$

注意到$\lambda_i(A) + \lambda_j(B) \leqslant z^*Az + z^*Bz$对于$V$中的任意单位向量$z$均成立, 对于$V$中的标准正交向量$z_1$, z_2, $\cdots$, z_{i+j-n}有

$$\begin{aligned}\lambda_i(A) + \lambda_j(B) &\leqslant \min_{\substack{z\in\mathcal{R}\left(\begin{bmatrix} z_1 & \cdots & z_{i+j-n}\end{bmatrix}\right)\\ \|z\|_2=1}} z^*(A+B)z \\ &\leqslant \max_{G\in\mathbb{C}_{i+j-n}^{n\times(i+j-n)}} \min_{\substack{z\in\mathcal{R}(G)\\ \|z\|_2=1}} z^*(A+B)z \\ &= \lambda_{i+j-n}(A+B).\end{aligned}$$

(ii) 类似于(i). □

推论 8.1.3 (Weyl**不等式**) 设$A,\ B\in\mathbb{C}^{n\times n}$满足$A^*=A,\ B^*=B$, A有特征值$\lambda_1(A)\geqslant\lambda_2(A)\geqslant\cdots\geqslant\lambda_n(A)$, 则

$$\lambda_i(A)+\lambda_{\max}(B)\geqslant\lambda_i(A+B)\geqslant\lambda_i(A)+\lambda_{\min}(B),\ i=1,2,\cdots,n.$$

证明 由定理8.1.6易见. □

推论 8.1.4 设$A,B\in\mathbb{C}^{n\times n}$满足$A\geqslant 0,\ B\geqslant 0,\ A-B\geqslant 0,\ \lambda_1(A)\geqslant\lambda_2(A)\geqslant\cdots\geqslant\lambda_n(A)$是$A$的特征值, $\lambda_1(B)\geqslant\lambda_2(B)\geqslant\cdots\geqslant\lambda_n(B)$是$B$的特征值, 则

$$\lambda_i(A)\geqslant\lambda_i(B),\ i=1,2,\cdots,n.$$

证明 由$A-B\geqslant 0$, 所以$\lambda_{\min}(A-B)\geqslant 0$. 故由推论8.1.3得

$$\lambda_i(A)=\lambda_i\left(B+(A-B)\right)\geqslant\lambda_i(B)+\lambda_{\min}(A-B)\geqslant\lambda_i(B),\ i=1,2,\cdots,n.$$

□

引理 8.1.1 设$A\in\mathbb{C}^{n\times n}$有奇异值$\sigma_1(A)\geqslant\cdots\geqslant\sigma_n(A)$, 则$\begin{bmatrix}0 & A^*\\ A & 0\end{bmatrix}$有特征值$\pm\sigma_1(A),\cdots,\pm\sigma_n(A)$.

证明 首先对A进行奇异值分解, 然后通过直接计算可得结论. □

推论 8.1.5 设$A,B\in\mathbb{C}^{n\times n}$, A有奇异值$\sigma_1(A)\geqslant\cdots\geqslant\sigma_n(A)$, 则

$$\sigma_i(A)+\sigma_{\max}(B)\geqslant\sigma_i(A+B)\geqslant\sigma_i(A)+\sigma_{\min}(B),\ i=1,\cdots,n.\tag{8.1.2}$$

Weyl不等式给出了两个矩阵的特征值与它们的和的特征值的关系,下面将证明两个半正定矩阵的特征值与它们的积的特征值之间也有类似的关系.

定理 8.1.7 设$A,\ B\in\mathbb{C}^{n\times n}$, A有奇异值$\sigma_1(A)\geqslant\cdots\geqslant\sigma_n(A)$, B有奇异值$\sigma_1(B)\geqslant\cdots\geqslant\sigma_n(B)$, 正整数$i,j$满足$i\leqslant n$且$j\leqslant n$, 则

(i) 当$i+j\geqslant n+1$时有$\sigma_{i+j-n}(AB)\geqslant\sigma_i(A)\sigma_j(B)$;

(ii) 当$i+j\leqslant n+1$时有$\sigma_{i+j-1}(AB)\leqslant\sigma_i(A)\sigma_j(B)$.

证明 由定理8.1.4知存在$C \in \mathbb{C}_i^{n\times i}$及$D \in \mathbb{C}_j^{n\times j}$使得

$$\sigma_i^2(A) = \min_{\substack{x\in\mathcal{R}(C)\\ \|x\|_2=1}} x^*A^*Ax, \quad \sigma_j^2(B) = \min_{\substack{y\in\mathcal{R}(D)\\ \|y\|_2=1}} y^*B^*By.$$

设$x_{i+1}, \cdots, x_n$是$\mathcal{R}(C)^\perp$的标准正交基, $Q = \mathcal{R}(B^*\begin{bmatrix}x_{i+1} & \cdots & x_n\end{bmatrix})^\perp$, 则

$$\begin{aligned}\dim Q = &\ n - \dim\mathcal{R}(B^*\begin{bmatrix}x_{i+1} & \cdots & x_n\end{bmatrix})\\ = &\ n - \operatorname{rank}(B^*\begin{bmatrix}x_{i+1} & \cdots & x_n\end{bmatrix})\\ \geqslant &\ n-(n-i)\\ = &\ i,\end{aligned}$$

于是

$$\dim(Q\cap\mathcal{R}(D)) = \dim Q + \dim\mathcal{R}(D) - \dim(Q+\mathcal{R}(D)) \geqslant i+j-n.$$

取$Q\cap\mathcal{R}(D)$中的标准正交向量$z_1, \cdots, z_{i+j-n}$, 则

$$z_k \in \mathcal{R}(D),\ \begin{bmatrix}x_{i+1} & \cdots & x_n\end{bmatrix}^* Bz_k = 0,\ k = 1,\cdots,i+j-n,$$

于是

$$Bz_k \in \mathcal{R}(C),\ k = 1,\cdots,i+j-n.$$

任取单位向量$z \in \mathcal{R}(\begin{bmatrix}z_1 & \cdots & z_{i+j-n}\end{bmatrix})$.

若$Bz=0$, 则$\sigma_j(B)^2 \leqslant z^*B^*Bz = 0$, 从而

$$\sigma_i(A)^2\sigma_j(B)^2 = 0 \leqslant \sigma_{i+j-n}(AB)^2.$$

若$Bz \neq 0$, 则

$$\sigma_i(A)^2\sigma_j(B)^2 \leqslant \left(\frac{Bz}{\|Bz\|_2}\right)^* A^*A\frac{Bz}{\|Bz\|_2} z^*B^*Bz = z^*B^*A^*Az,$$

于是

$$\begin{aligned}\sigma_i(A)^2\sigma_j(B)^2 &\leqslant \min_{z\in\mathcal{R}(\begin{bmatrix}z_1 & \cdots & z_{i+j-n}\end{bmatrix})} z^*B^*A^*Az\\ &\leqslant \max_{G\in\mathbb{C}_{i+j-n}^{n\times(i+j-n)}} \min_{\substack{z\in\mathcal{R}(G)\\ \|z\|_2=1}} z^*B^*A^*Az = \sigma_{i+j-n}(AB)^2,\end{aligned}$$

因而$\sigma_i(A)\sigma_j(B) \leqslant \sigma_{i+j-n}(AB)$.

(ii) 类似于(i). □

推论 8.1.6 设$A, B \in \mathbb{C}^{n\times n}$, A有奇异值$\sigma_1(A) \geqslant \cdots \geqslant \sigma_n(A)$, 则

$$\sigma_i(A)\sigma_{\max}(B) \geqslant \sigma_i(AB) \geqslant \sigma_i(A)\sigma_{\min}(B),\ i = 1,\cdots,n. \tag{8.1.3}$$

推论 8.1.7 设A, $B \in \mathbb{C}^{n\times n}$, $A \geqslant 0$, $B \geqslant 0$, A有特征值$\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$, 则

$$\lambda_i(A)\lambda_{\max}(B) \geqslant \lambda_i(AB) \geqslant \lambda_i(A)\lambda_{\min}(B).$$

证明 注意到

$$\lambda_i(AB) = \lambda_i\left((B^{\frac{1}{2}})^*(A^{\frac{1}{2}})^*A^{\frac{1}{2}}B^{\frac{1}{2}}\right) = \sigma_i\left(A^{\frac{1}{2}}B^{\frac{1}{2}}\right)^2$$

及

$$\lambda_i(A) = \lambda_i\left((A^{\frac{1}{2}})^*A^{\frac{1}{2}}\right) = \sigma_i(A^{\frac{1}{2}})^2,$$

对$A^{\frac{1}{2}}$及$B^{\frac{1}{2}}$应用推论8.1.6得证. □

8.2 Kantorarich不等式

定理 8.2.1 (Kantorarich不等式) 设$A \in \mathbb{C}^{n\times n}$满足$A > 0$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$是$A$的特征值, 则

$$1 \leqslant \frac{(x^*Ax)(x^*A^{-1}x)}{(x^*x)^2} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}, \quad \forall x \in \mathbb{C}^n,\ x \neq 0.$$

证明 先证$1 \leqslant \frac{(x^*Ax)(x^*A^{-1}x)}{(x^*x)^2}$. 由Cauchy-Schwarz不等式知

$$x^*x = \left(x^*A^{\frac{1}{2}}\right)\left(A^{-\frac{1}{2}}x\right) \leqslant \sqrt{x^*Ax}\cdot\sqrt{x^*A^{-1}x},$$

从而$(x^*x)^2 \leqslant (x^*Ax)\cdot(x^*A^{-1}x)$, 即

$$1 \leqslant \frac{(x^*Ax)(x^*A^{-1}x)}{(x^*x)^2}.$$

再证$\frac{(x^*Ax)(x^*A^{-1}x)}{(x^*x)^2} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}$. 注意到

$$\begin{aligned}
&\frac{(x^*Ax)(x^*A^{-1}x)}{(x^*x)^2} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n} \\
\Leftrightarrow\quad & (x^*Ax)(x^*A^{-1}x) \leqslant (x^*x)^2\frac{\lambda_1+\lambda_n}{2}\cdot\frac{\lambda_1^{-1}+\lambda_n^{-1}}{2} \\
\Leftrightarrow\quad & 4\frac{x^*Ax}{\lambda_1+\lambda_n}\cdot\frac{x^*A^{-1}x}{\lambda_1^{-1}+\lambda_n^{-1}} \leqslant (x^*x)^2 \\
\Leftrightarrow\quad & 2\sqrt{\frac{x^*Ax}{\lambda_1+\lambda_n}}\cdot\sqrt{\frac{x^*A^{-1}x}{\lambda_1^{-1}+\lambda_n^{-1}}} \leqslant x^*x.
\end{aligned}$$

由$A>0$知存在酉矩阵Q使得

$$A=Q\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)Q^*.$$

设$y=Q^*x=[y_1\ y_2\ \cdots\ y_n]^*$, 则

$$\begin{aligned}
&\frac{(x^*Ax)(x^*A^{-1}x)}{(x^*x)^2}\leqslant\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}\\
\Leftrightarrow\quad&2\sqrt{\frac{y^*\mathrm{diag}(\lambda_1\ \cdots\ \lambda_n)y}{\lambda_1+\lambda_n}}\cdot\sqrt{\frac{y^*\mathrm{diag}(\lambda_1^{-1}\ \cdots\ \lambda_n^{-1})y}{\lambda_1^{-1}+\lambda_n^{-1}}}\leqslant y^*y\\
\Leftrightarrow\quad&2\sqrt{\frac{\lambda_1y_1^2+\lambda_2y_2^2+\cdots+\lambda_ny_n^2}{\lambda_1+\lambda_n}}\cdot\sqrt{\frac{\lambda_1^{-1}y_1^2+\lambda_2^{-1}y_2^2+\cdots+\lambda_n^{-1}y_n^2}{\lambda_1^{-1}+\lambda_n^{-1}}}\\
&\leqslant y_1^2+y_2^2+\cdots+y_n^2.
\end{aligned}$$

由算术平均不等式, 只需证

$$\frac{\lambda_1y_1^2+\lambda_2y_2^2+\cdots+\lambda_ny_n^2}{\lambda_1+\lambda_n}+\frac{\lambda_1^{-1}y_1^2+\lambda_2^{-1}y_2^2+\cdots+\lambda_n^{-1}y_n^2}{\lambda_1^{-1}+\lambda_n^{-1}}\leqslant y_1^2+y_2^2+\cdots+y_n^2.$$

只需证

$$\frac{\lambda_i}{\lambda_1+\lambda_n}+\frac{\lambda_i^{-1}}{\lambda_1^{-1}+\lambda_n^{-1}}\leqslant 1.$$

事实上, 由$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n>0$及$\lambda_1^{-1}\leqslant\lambda_2^{-1}\leqslant\cdots\leqslant\lambda_n^{-1}$得

$$(\lambda_i-\lambda_n)\left(\lambda_1^{-1}-\lambda_i^{-1}\right)+(\lambda_i-\lambda_1)\left(\lambda_n^{-1}-\lambda_i^{-1}\right)\leqslant 0,$$

即

$$(\lambda_i-\lambda_n)\lambda_1^{-1}-(\lambda_i-\lambda_n)\lambda_i^{-1}+(\lambda_i-\lambda_1)\lambda_n^{-1}-(\lambda_i-\lambda_1)\lambda_i^{-1}\leqslant 0,$$

亦即

$$\lambda_i\lambda_1^{-1}+\lambda_n^{-1}\lambda_i+\lambda_i^{-1}\lambda_1+\lambda_i^{-1}\lambda_n\leqslant 2+\lambda_1\lambda_n^{-1}+\lambda_n\lambda_1^{-1},$$

从而

$$\lambda_i\left(\lambda_1^{-1}+\lambda_n^{-1}\right)+\lambda_i^{-1}(\lambda_1+\lambda_n)\leqslant(\lambda_1+\lambda_n)\left(\lambda_1^{-1}+\lambda_n^{-1}\right),$$

故

$$\frac{\lambda_i}{\lambda_1+\lambda_n}+\frac{\lambda_i^{-1}}{\lambda_1^{-1}+\lambda_n^{-1}}\leqslant 1.$$

□

8.3 Courant-Fischer定理的推广

本节推广定理8.1.4到更一般的情形.

设$H \in \mathbb{C}^{n\times n}$满足$H^* = H$, $\varphi(t_1,\cdots,t_k)$是定义在某个区域上的k元实值对称函数, 假定$\varphi(t_1,\cdots,t_k)$关于每个变量均为增函数.

定义函数$\widetilde{\varphi}:\mathbb{C}^n\times\cdots\times\mathbb{C}^n\to\mathbb{R}$如下:

$$\widetilde{\varphi}(x_1,\cdots,x_k)=\varphi(\mu_1,\cdots,\mu_k),\ \forall(x_1,\cdots,x_k)\in\mathbb{C}^n\times\cdots\times\mathbb{C}^n,$$

其中$\mu_1,\cdots,\mu_k$是矩阵$[x_1\cdots x_k]^*H[x_1\cdots x_k]$的特征值.

对于正整数$i_1,\cdots,i_k$及$j_1,\cdots,j_k$, 用符号

$$(j_1,j_2,\cdots,j_k)\leqslant(i_1,i_2,\cdots,i_k)$$

表示$j_p\leqslant i_p, p=1,2,\cdots,k$. 若正整数$i_1,\cdots,i_k$满足$i_1<i_2<\cdots<i_k(i_1\leqslant i_2\leqslant\cdots\leqslant i_k)$, 称$(i_1,\cdots,i_k)$是严格递增的(非降的).

定义 8.3.1 对于每个满足$i_p\geqslant p, p=1,2,\cdots,k$的非降正整数序列$i_1$, i_2, $\cdots$, i_k, 用i_1', i_2', $\cdots$, i_k'定义满足下面(i)和(ii)的严格递增的正整数序列.

(i) $(i_1',i_2',\cdots,i_k')\leqslant(i_1,i_2,\cdots,i_k)$;

(ii) 若严格递增的正整数序列$(j_1,\cdots,j_k)$满足$(j_1,\cdots,j_k)\leqslant(i_1,\cdots,i_k)$, 则

$$(j_1,\cdots,j_k)\leqslant(i_1',i_2',\cdots,i_k').$$

清楚地, $(i_1',i_2',\cdots,i_k')$是不大于非降正整数序列$(i_1,i_2,\cdots,i_k)$的最大的严格递增的正整数序列. 例如: 当$(i_1,i_2,\cdots,i_k)$严格递增时, 则$i_p'=i_p, p=1,2,\cdots,k$; 当$i_1=i_2=\cdots=i_k$且$i_k\geqslant k$时, 则$(i_1',i_2',\cdots,i_k')=(i_k-k+1,\cdots,i_k-1,i_k)$.

类似地, 对于每个满足$i_p\leqslant n-k+p, p=1,2,\cdots,k$的非降正整数序列$i_1,i_2,\cdots,i_k$, 用$(i_1'',i_2'',\cdots,i_k'')$定义不小于非降正整数序列$(i_1,i_2,\cdots,i_k)$的最小的严格递增的正整数序列. 显然, 若$(i_1,i_2,\cdots,i_k)$ 是一个严格递增的正整数序列, 则$i_s'=i_s=i_s''$, $s=1,2,\cdots,k$.

命题 8.3.1 若正整数$i_1,\cdots,i_k$满足$i_1\leqslant\cdots\leqslant i_{k-1}<i_k$且

$$i_p\geqslant p,\ p=1,2,\cdots,k,$$

则$i_k'=i_k$.

证明 由定义8.3.1知$i_k^{'} \leqslant i_k$. 若$i_k^{'} < i_k$, 则

$$(i_1^{'},\cdots,i_{k-1}^{'},i_k) \leqslant (i_1,\cdots,i_k)$$

且

$$(i_1^{'},\cdots,i_k^{'}) \leqslant (i_1^{'},\cdots,i_{k-1}^{'},i_k),$$

与定义8.3.1矛盾. □

推论 8.3.1 设正整数$i_1,\cdots,i_k$满足$i_1 \leqslant \cdots \leqslant i_k$且$i_p \geqslant p$, $p=1,\cdots,k$, 定义$\mathcal{L}(i_1,\cdots,i_k)=(i_1^{'},\cdots,i_k^{'})$. 则

(i) 当$i_{k-1} < i_k$时有$\mathcal{L}(i_1,\cdots,i_{k-1})=(i_1^{'},\cdots,i_{k-1}^{'})$且$i_k^{'}=i_k$;

(ii) 当$i_1 \leqslant \cdots \leqslant i_m < i_{m+1}=\cdots=i_k$时有

$$\mathcal{L}(i_1,\cdots,i_m)=(i_1^{'},\cdots,i_m^{'})\text{且}i_p^{'}=i_k-(k-p),\ p=m+1,\cdots,k.$$

证明 由定义8.3.1及命题8.3.1易见. □

类似于命题8.3.1, 我们有

命题 8.3.2 若正整数$i_1,\cdots,i_k$满足$i_1 < i_2 \leqslant \cdots \leqslant i_k$且$i_p \leqslant n-k+p$, $p=1,2,\cdots,k$. 则$i_1^{''}=i_1$.

推论 8.3.2 设正整数$i_1,\cdots,i_k$满足$i_1 \leqslant \cdots \leqslant i_k$且$i_p \leqslant n-k+p$, $p=1,2,\cdots,k$, 用$\widetilde{\mathcal{L}}(i_1,\cdots,i_k)$记$(i_1^{''},\cdots,i_k^{''})$, 则

(i) 当$i_1 < i_2$时有$i_1^{''}=i_1$且$(i_2^{''},\cdots,i_k^{''})=\widetilde{\mathcal{L}}(i_2,\cdots,i_k)$;

(ii) 当$i_1=\cdots=i_{m-1}=i_m < i_{m+1} \leqslant \cdots \leqslant i_k$时有$i_p^{''}=i_1+p-1$, $p=1,2,\cdots,m$且$\widetilde{\mathcal{L}}(i_{m+1},\cdots,i_k)=(i_{m+1}^{''},\cdots,i_k^{''})$.

证明 由命题8.3.2易见. □

命题 8.3.3 若非降的正整数序列$(i_1,i_2,\cdots,i_k)$和$(j_1,j_2,\cdots,j_k)$满足

$$i_p \geqslant p, j_p \geqslant p, p=1,2,\cdots,k,$$

则

$$\left((i_1+j_1)',(i_2+j_2)',\cdots,(i_k+j_k)'\right) \geqslant \left(i_1'+j_1',i_2'+j_2',\cdots,i_k'+j_k'\right).$$

证明 对于$p=1,2,\cdots,k$, 易见$i'_p+j'_p\leqslant i_p+j_p$, 即

$$(i'_1+j'_1,i'_2+j'_2,\cdots,i'_k+j'_k)\leqslant(i_1+j_1,i_2+j_2,\cdots,i_p+j_p).$$

由定义8.3.1, 结论成立. □

定理 8.3.1 设H, φ及$\widetilde{\varphi}$同前面定义, 正整数i_1, $\cdots$, i_k满足$i_1\leqslant\cdots\leqslant i_k\leqslant n$且$i_p\geqslant p$, $p=1,2,\cdots,k$, 则

$$\sup_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p=i_p}}\inf_{\substack{x_p\in M_p\\ \{x_p\}o.n.}}\widetilde{\varphi}(x_1,\cdots,x_k)=\varphi(\lambda_{i'_1},\cdots,\lambda_{i'_k}).$$

证明 证明见附录E. □

推论 8.3.3 设H, φ及$\widetilde{\varphi}$同前面定义, 正整数$i_1,\cdots,i_k$满足$1\leqslant i_1<i_2<\cdots<i_k\leqslant n$, 则

$$\sup_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p=i_p}}\inf_{\substack{x_p\in M_p\\ \{x_p\}o.n.}}\widetilde{\varphi}(x_1,\cdots,x_k)=\varphi(\lambda_{i_1},\cdots,\lambda_{i_k}).$$

定理 8.3.2 设H, φ及$\widetilde{\varphi}$同前面定义, 正整数$i_1,\cdots,i_k$满足$i_1\leqslant\cdots\leqslant i_k$, 且$i_p\leqslant n-k+p$, $p=1,2,\cdots,k$, $k\leqslant n$, 则

$$\inf_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p=i_p-1}}\sup_{\substack{x_p\perp M_p\\ \{x_p\}o.n.}}\widetilde{\varphi}(x_1,\cdots,x_k)=\varphi(\lambda_{i''_1},\cdots,\lambda_{i''_k}),$$

其中$(i''_1,\cdots,i''_k)$是大于或等于$(i_1,\cdots,i_k)$的严格递增的最小的正整数序列.

证明 令$\psi(t_1,\cdots,t_k)=-\varphi(-t_1,\cdots,-t_k)$, 则$\psi$是对称的且关于每个变元均为增函数.

设$j_p=n+1-i_{k+1-p}$, $p=1,2,\cdots,k$, 则$1\leqslant j_1\leqslant\cdots\leqslant j_k\leqslant n$且$j_p\geqslant p$, $p=1,2,\cdots,k$. 对于$-H$及ψ使用定理8.3.1推出

$$\begin{aligned}\sup_{\substack{N_1\subseteq\cdots\subseteq N_k\\ \dim N_p=j_p}}\inf_{\substack{x_p\in N_p\\ \{x_p\}o.n.}}\widetilde{\psi}(x_1,\cdots,x_k)&=\psi(-\lambda_{n+1-j'_1},\cdots,-\lambda_{n+1-j'_k})\\ &=-\varphi(\lambda_{n+1-j'_1},\cdots,\lambda_{n+1-j'_k}),\end{aligned}$$

使用

$$\widetilde{\psi}(x_1,\cdots,x_k)=-\widetilde{\varphi}(x_1,\cdots,x_k)$$

及

$$(j'_1,\cdots,j'_k)=(n+1-i''_k,\cdots,n+1-i''_1),$$

并令$M_p=N^{\perp}_{k+1-p}$, 推出结论成立. □

推论 8.3.4 设正整数$i_1,\cdots,i_k$满足$i_1<\cdots<i_k$, 且$i_p\leqslant n-k+p,p=1,2,\cdots,k$, H, φ及$\widetilde{\varphi}$同前面定义, 则

$$\inf_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p=i_p-1}}\sup_{\substack{x_p\perp M_p\\ \{x_p\}o.n.}}\widetilde{\varphi}(x_1,\cdots,x_k)=\varphi(\lambda_{i_1},\cdots,\lambda_{i_k}).$$

推论 8.3.5 设$H\in\mathbb{C}^{n\times n}$满足$H\geqslant 0$, $\lambda_1\geqslant\cdots\geqslant\lambda_n$是$H$的特征值, 正整数$i_1,\cdots,i_k$满足$i_1\leqslant\cdots\leqslant i_k\leqslant n$且$i_p\geqslant p,p=1,2,\cdots,k$, 则

$$\sup_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p=i_p}}\inf_{\substack{x_p\in M_p\\ \{x_p\}o.n.}}\det\left([x_1\cdots x_k]^*H[x_1\cdots x_k]\right)=\lambda_{i_1'}\cdots\lambda_{i_k'}.$$

证明 令φ同前面定义, 既然$\det\left([x_1\cdots x_k]^*H[x_1\cdots x_k]\right)=\mu_1\cdots\mu_k$, 在定理8.3.1中取$\varphi(t_1,\cdots,t_k)=t_1\cdots t_k$即可. □

推论 8.3.6 设$H\in\mathbb{C}^{n\times n}$满足$H\geqslant 0$, $\lambda_1\geqslant\cdots\geqslant\lambda_n$是$H$的特征值, 正整数$i_1,\cdots,i_k$满足$i_1<\cdots<i_k\leqslant n$, 则

$$\sup_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p=i_p}}\inf_{\substack{x_p\in M_p\\ \{x_p\}o.n.}}\det\left([x_1\cdots x_k]^*H[x_1\cdots x_k]\right)=\lambda_{i_1}\cdots\lambda_{i_k}.$$

证明 由推论8.3.5易见. □

推论 8.3.7 设$H\in\mathbb{C}^{n\times n}$满足$H\geqslant 0$, $\lambda_1\geqslant\cdots\geqslant\lambda_n$是$H$的特征值, 正整数$i_1,\cdots,i_k$满足$i_1\leqslant\cdots\leqslant i_k$且$i_p\leqslant n-k+p,p=1,2,\cdots,k$, $k\leqslant n$, 则

$$\inf_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p=i_p-1}}\sup_{\substack{x_p\perp M_p\\ \{x_p\}o.n.}}\det\left([x_1\cdots x_k]^*H[x_1\cdots x_k]\right)=\lambda_{i_1''}\cdots\lambda_{i_k''}.$$

证明 在定理8.3.2中取$\varphi(t_1,\cdots,t_k)=t_1\cdots t_k$即可. □

推论 8.3.8 设$H\in\mathbb{C}^{n\times n}$满足$H\geqslant 0$, $\lambda_1\geqslant\cdots\geqslant\lambda_n$是$H$的特征值, 正整数$i_1,\cdots,i_k$满足$i_1<\cdots<i_k\leqslant n$且$i_p\leqslant n-k+p,p=1,2,\cdots,k$, 则

$$\inf_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p=i_p-1}}\sup_{\substack{x_p\perp M_p\\ \{x_p\}o.n.}}\det\left([x_1\cdots x_k]^*H[x_1\cdots x_k]\right)=\lambda_{i_1}\cdots\lambda_{i_k}.$$

证明 由推论8.3.7易见. □

推论 8.3.9 设$H \in \mathbb{C}^{n\times n}$满足$H^* = H$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$是$H$的特征值, 正整数$i_1, \cdots, i_k$满足$i_1 \leqslant \cdots \leqslant i_k \leqslant n$且$i_p \geqslant p, p = 1, 2, \cdots, k$, 则

$$\sup_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p = i_p}} \inf_{\substack{x_p\in M_p\\ \{x_p\} o.n.}} (x_1^* H x_1 + \cdots + x_k^* H x_k) = \lambda_{i_1'} + \cdots + \lambda_{i_k'}.$$

证明 令$\tilde{\varphi}$同前面定义, 注意到$\mu_1 + \cdots + \mu_k = x_1^* H x_1 + \cdots + x_k^* H x_k$, 在定理8.3.1中取$\varphi(t_1, \cdots, t_k) = t_1 + \cdots + t_k$即可. □

推论 8.3.10 设$H \in \mathbb{C}^{n\times n}$满足$H^* = H$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$是$H$的特征值, 正整数$i_1, \cdots, i_k$满足$i_1 < \cdots < i_k \leqslant n$, 则

$$\sup_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p = i_p}} \inf_{\substack{x_p\in M_p\\ \{x_p\} o.n.}} (x_1^* H x_1 + \cdots + x_k^* H x_k) = \lambda_{i_1} + \cdots + \lambda_{i_k}.$$

证明 由推论8.3.9易见. □

推论 8.3.11 设$H \in \mathbb{C}^{n\times n}$满足$H^* = H$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$是$H$的特征值, 正整数$i_1, \cdots, i_k$满足$i_1 \leqslant \cdots \leqslant i_k$且$i_p \leqslant n - k + p$, $p = 1, 2, \cdots, k$, $k \leqslant n$, 则

$$\inf_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p = i_p - 1}} \sup_{\substack{x_p\perp M_p\\ \{x_p\} o.n.}} (x_1^* H x_1 + \cdots + x_k^* H x_k) = \lambda_{i_1''} + \cdots + \lambda_{i_k''}.$$

证明 在定理8.3.2中取$\varphi(t_1, \cdots, t_k) = t_1 + \cdots + t_k$即可. □

推论 8.3.12 设$H \in \mathbb{C}^{n\times n}$满足$H^* = H$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$是$H$的特征值, 正整数$i_1, \cdots, i_k$满足$i_1 < \cdots < i_k \leqslant n$且$i_p \leqslant n - k + p$, $p = 1, 2, \cdots, k$, 则

$$\inf_{\substack{M_1\subseteq\cdots\subseteq M_k\\ \dim M_p = i_p - 1}} \sup_{\substack{x_p\perp M_p\\ \{x_p\} o.n.}} (x_1^* H x_1 + \cdots + x_k^* H x_k) = \lambda_{i_1} + \cdots + \lambda_{i_k}.$$

证明 由推论8.3.11易见. □

8.4 两个矩阵乘积的奇异值和特征值不等式

本节给出一些关于两个矩阵乘积的奇异值和特征值不等式.

引理 8.4.1 设$H \in \mathbb{C}^{n\times n}$, $H^* = H$, $y_1, \cdots, y_k \in \mathbb{C}^n$线性无关, $u_1, \cdots, u_k$是$\mathrm{span}\{y_1, \cdots, y_k\}$的标准正交基, 则

$$\det([y_1 \cdots y_k]^* H[y_1 \cdots y_k]) = \det([u_1 \cdots u_k]^* H[u_1 \cdots u_k]) \det[(y_i, y_j)].$$

证明 由$y_1, \cdots, y_k \in \mathbb{C}^n$线性无关及Schmidt正交化过程知, 存在上三角阵C使得

$$[y_1 \cdots y_k] = [u_1 \cdots u_k]C,$$

于是

$$\begin{aligned}
& \det([y_1 \cdots y_k]^* H[y_1 \cdots y_k]) \\
= \ & \det(C^*[u_1 \cdots u_k]^* H[u_1 \cdots u_k]C) \\
= \ & \det(C^*C)\det([u_1 \cdots u_k]^* H[u_1 \cdots u_k]) \\
= \ & \det(C^*[u_1 \cdots u_k]^*[u_1 \cdots u_k]C)\det([u_1 \cdots u_k]^* H[u_1 \cdots u_k]) \\
= \ & \det[(y_i, y_j)]\det([u_1 \cdots u_k]^* H[u_1 \cdots u_k]).
\end{aligned}$$

□

定理 8.4.1 设A, $B \in \mathbb{C}^{n\times n}$, $\sigma_1(A) \geqslant \cdots \geqslant \sigma_n(A)$是$A$的奇异值, $\sigma_1(B) \geqslant \cdots \geqslant \sigma_n(B)$是$B$的奇异值, $\sigma_1(AB) \geqslant \cdots \geqslant \sigma_n(AB)$是$AB$的奇异值, $1 \leqslant i_1 \leqslant \cdots \leqslant i_k \leqslant n$, $1 \leqslant j_1 \leqslant \cdots \leqslant j_k \leqslant n$, 序列$i_1', \cdots, i_k'$; $i_1'', \cdots, i_k''$; $j_1', \cdots, j_k'$和$j_1'', \cdots, j_k''$同上节中定义, 则

(i) 当$i_p \geqslant p, j_p \geqslant p, i_p + j_p \geqslant n + p, p = 1, 2, \cdots, k$时有

$$\begin{aligned}
& \sigma_{i_1'}(A) \cdots \sigma_{i_k'}(A)\sigma_{j_1'}(B) \cdots \sigma_{j_k'}(B) \\
\leq \ & \sigma_{(i_1+j_1-n)'}(AB) \cdots \sigma_{(i_k+j_k-n)'}(AB);
\end{aligned}$$

(ii) 当$i_p + j_p \leqslant n - k + p + 1, p = 1, 2, \cdots, k$时有

$$\begin{aligned}
& \sigma_{i_1''}(A) \cdots \sigma_{i_k''}(A)\sigma_{j_1''}(B) \cdots \sigma_{j_k''}(B) \\
\geqslant \ & \sigma_{(i_1+j_1-1)''}(AB) \cdots \sigma_{(i_k+j_k-1)''}(AB).
\end{aligned}$$

证明 (i) 由推论8.3.5知存在$\mathbb{C}^n$的子空间$M_1 \subseteq \cdots \subseteq M_k$及$N_1 \subseteq \cdots \subseteq N_k$使得$\dim M_p = i_p$, $\dim N_p = j_p$, $p = 1, \cdots, k$, 并且

$$\inf_{\substack{x_p \in M_p \\ \{x_p\} o.n.}} \det([x_1 \cdots x_k]^* A^* A[x_1 \cdots x_k]) = \sigma_{i_1'}(A)^2 \cdots \sigma_{i_k'}(A)^2, \qquad (8.4.1)$$

$$\inf_{\substack{x_p\in N_p\\ \{x_p\}o.n.}} \det\left([x_1\cdots x_k]^*B^*B[x_1\cdots x_k]\right)=\sigma_{j_1'}(B)^2\cdots\sigma_{j_k'}(B)^2. \qquad (8.4.2)$$

令$Q_p=[\mathcal{R}(B^*M_p^{\perp})]^{\perp}$, $p=1,2,\cdots,k$, 则

$$\dim Q_p=n-\dim\mathcal{R}(B^*M_p^{\perp})\geqslant n-\dim M_p^{\perp}=i_p,$$

于是

$$\dim(Q_p\cap N_p)=\dim Q_p+\dim N_p-\dim(Q_p+N_p)\geqslant i_p+j_p-n.$$

故存在$Q_k\cap N_k$的子空间$P_{i_1+j_1-n}\subseteq\cdots\subseteq P_{i_k+j_k-n}$使得$\dim P_{i_p+j_p-n}=i_p+j_p-n$, $p=1,2,\cdots,k$.

任取标准正交向量组$y_p\in P_{i_p+j_p-n}$, $p=1,\cdots,k$, 则$By_p\in M_p, y_p\in N_p$, $p=1,2,\cdots,k$. 若$By_1,\cdots,By_k$线性相关, 则由式(8.4.2)知

$$\sigma_{j_1'}(B_1)\cdots\sigma_{j_k'}(B_k)=0,$$

从而结论成立. 若$By_1,\cdots,By_k$线性无关, 则由引理8.4.1得

$$\begin{aligned}&\det\left([By_1\cdots By_k]^*A^*A[By_1\cdots By_k]\right)\\ =\ &\det\left([\mu_1\cdots\mu_k]^*A^*A[\mu_1\cdots\mu_k]\right)\cdot\det\left([By_1\cdots By_k]^*[By_1\cdots By_k]\right),\end{aligned} \qquad (8.4.3)$$

其中$\mu_1,\cdots,\mu_k$是$\mathrm{span}\{By_1,\cdots,By_k\}$的标准正交基, $\mu_p\in M_p$, $p=1,\cdots,k$. 结合式(8.4.1)、式(8.4.2)、式(8.4.3)得到

$$\begin{aligned}&\sigma^2_{i_1'}(A)\cdots\sigma^2_{i_k'}(A)\sigma^2_{j_1'}(B)\cdots\sigma^2_{j_k'}(B)\\ \leqslant\ &\det\left([By_1\cdots By_k]^*A^*A[By_1\cdots By_k]\right)\\ =\ &\det\left([y_1\cdots y_k]^*B^*A^*AB[y_1\cdots y_k]\right)\end{aligned}$$

使用$\{y_p\}$的任意性推出

$$\begin{aligned}&\inf_{\substack{y_p\in P_{i_p+j_p-n}\\ \{y_p\}o.n.}}\det\left([y_1\cdots y_k]^*B^*A^*AB[y_1\cdots y_k]\right)\\ \geqslant\ &\sigma^2_{i_1'}(A)\cdots\sigma^2_{i_k'}(A)\sigma^2_{j_1'}(B)\cdots\sigma^2_{j_k'}(B).\end{aligned}$$

再由推论8.3.5得证.

(ii) 使用推论8.3.7可类似于(i)证得结论(留给读者). □

推论 8.4.1 设A, $B\in\mathbb{C}^{n\times n}$, A有奇异值$\sigma_1(A)\geqslant\cdots\geqslant\sigma_n(A)$, B有奇异值$\sigma_1(B)\geqslant\cdots\geqslant\sigma_n(B)$, AB有奇异值$\sigma_1(AB)\geqslant\cdots\geqslant\sigma_n(AB)$, 正整数$i_1,\cdots,i_k$满足$i_1<\cdots<i_k<n$, 则

(i) $\sigma_{i_1}(A)\cdots\sigma_{i_k}(A)\sigma_1(B)\cdots\sigma_k(B)\geqslant\sigma_{i_1}(AB)\cdots\sigma_{i_k}(AB)$;

(ii) $\sigma_1(A)\cdots\sigma_k(A)\sigma_{i_1}(B)\cdots\sigma_{i_k}(B)\geqslant\sigma_{i_1}(AB)\cdots\sigma_{i_k}(AB)$;

(iii) $\sigma_{i_1}(A)\cdots\sigma_{i_k}(A)\sigma_{n-k+1}(B)\cdots\sigma_n(B)\leqslant\sigma_{i_1}(AB)\cdots\sigma_{i_k}(AB)$;

(iv) $\sigma_{n-k+1}(A)\cdots\sigma_n(A)\sigma_{i_1}(B)\cdots\sigma_{i_k}(B)\leqslant\sigma_{i_1}(AB)\cdots\sigma_{i_k}(AB)$.

证明 (i) 在定理8.4.1(ii)中取$j_1=\cdots=j_k=1$得到

$$\sigma_{i_1''}(A)\cdots\sigma_{i_k''}(A)\sigma_{1''}(B)\cdots\sigma_{1''}(B)\geqslant\sigma_{i_1''}(AB)\cdots\sigma_{i_k''}(AB).$$

这和$i_p''=i_p,p=1,2,\cdots,k$一起推出结论(i)成立.

(ii)~(iv)的证明类似于(i). □

推论 8.4.2 设A, $B\in\mathbb{C}^{n\times n}$, $A\geqslant 0$, $B\geqslant 0$, A有特征值$\lambda_1(A)\geqslant\cdots\geqslant\lambda_n(A)$, B有特征值$\lambda_1(B)\geqslant\cdots\geqslant\lambda_n(B)$, AB有特征值$\lambda_1(AB)\geqslant\cdots\geqslant\lambda_n(AB)$, 正整数$i_1,\cdots,i_k$ 满足$i_1<\cdots<i_k\leqslant n$, 则

(i) $\lambda_{i_1}(A)\cdots\lambda_{i_k}(A)\lambda_1(B)\cdots\lambda_k(B)\geqslant\lambda_{i_1}(AB)\cdots\lambda_{i_k}(AB)$;

(ii) $\lambda_1(A)\cdots\lambda_k(A)\lambda_{i_1}(B)\cdots\lambda_{i_k}(B)\geqslant\lambda_{i_1}(AB)\cdots\lambda_{i_k}(AB)$;

(iii) $\lambda_{i_1}(A)\cdots\lambda_{i_k}(A)\lambda_{n-k+1}(B)\cdots\lambda_n(B)\leqslant\lambda_{i_1}(AB)\cdots\lambda_{i_k}(AB)$;

(iv) $\lambda_{n-k+1}(A)\cdots\lambda_n(A)\lambda_{i_1}(B)\cdots\lambda_{i_k}(B)\leqslant\lambda_{i_1}(AB)\cdots\lambda_{i_k}(AB)$.

证明 注意到$\sigma_{i_p}(A^{\frac{1}{2}})=\sqrt{\lambda_{i_p}(A)}$, $\sigma_{i_p}(B^{\frac{1}{2}})=\sqrt{\lambda_{i_p}(B)}$, $\lambda_{i_p}(AB)=\lambda_{ip}(B^{\frac{1}{2}}AB^{\frac{1}{2}})=\sigma_{i_p}(A^{\frac{1}{2}}B^{\frac{1}{2}})$, $p=1,2,\cdots,k$. 由推论8.4.1得证. □

8.5 两个矩阵和的奇异值和特征值不等式

本节给出一些关于两个Hermite矩阵之和的奇异值和特征值不等式.

定理 8.5.1 设A, $B\in\mathbb{C}^{n\times n}$满足$A^*=A$且$B^*=B$, $\lambda_1(A)\geqslant\cdots\geqslant\lambda_n(A)$是$A$的特征值, $\lambda_1(B)\geqslant\cdots\geqslant\lambda_n(B)$是$B$的特征值, $\lambda_1(A+B)\geqslant\cdots\geqslant\lambda_n(A+B)$是$A+B$的特征值, 正整数$i_1,\cdots,i_k$满足$i_1\leqslant\cdots\leqslant i_k$, 正整数$j_1,\cdots,j_k$满足$j_1\leqslant\cdots\leqslant j_k$, 则

(i) 当$i_k \leqslant n$, $j_k \leqslant n$且$i_p + j_p \geqslant n + p$, $p = 1, 2, \cdots, k$时有

$$\begin{aligned} & \lambda_{i_1'}(A) + \cdots + \lambda_{i_k'}(A) + \lambda_{j_1'}(B) + \cdots + \lambda_{j_k'}(B) \\ \leqslant \ & \lambda_{(i_1+j_1-n)'}(A+B) + \cdots + \lambda_{(i_k+j_k-n)'}(A+B); \end{aligned}$$

(ii) 当$k \leqslant n$且$i_p + j_p \leqslant n - k + p + 1$, $p = 1, 2, \cdots, k$时有

$$\begin{aligned} & \lambda_{i_1''}(A) + \cdots + \lambda_{i_k''}(A) + \lambda_{j_1''}(B) + \cdots + \lambda_{j_k''}(B) \\ \geqslant \ & \lambda_{(i_1+j_1-1)''}(A+B) + \cdots + \lambda_{(i_k+j_k-1)''}(A+B). \end{aligned}$$

证明 (i) 由$i_p + j_p \geqslant n + p$, $i_p \leqslant n$及$j_p \leqslant n$得$i_p \geqslant p$且$j_p \geqslant p$, $p = 1, 2, \cdots, k$. 由推论8.3.9知存在$\mathbb{C}^n$的子空间$M_1, \cdots, M_k$及$N_1, \cdots, N_k$满足$M_1 \subseteq \cdots \subseteq M_k$, $N_1 \subseteq \cdots \subseteq N_k$, $\dim M_p = i_p$, $\dim N_p = j_p$, $p = 1, \cdots, k$, 并且

$$\inf_{\substack{x_p \in M_p \\ \{x_p\} o.n.}} (x_1^* A x_1 + \cdots x_k^* A x_k) = \lambda_{i_1'}(A) + \cdots + \lambda_{i_k'}(A), \tag{8.5.1}$$

$$\inf_{\substack{x_p \in N_p \\ \{x_p\} o.n.}} (x_1^* B x_1 + \cdots + x_k^* B x_k) = \lambda_{j_1'}(B) + \cdots + \lambda_{j_k'}(B). \tag{8.5.2}$$

既然$\dim(M_p \cap N_p) = \dim M_p + \dim N_p - \dim(M_p + N_p) \geqslant i_p + j_p - n$, $p = 1, 2, \cdots, k$, 存在$M_k \cap N_k$的子空间$P_{i_1+j_1-n}, \cdots, P_{i_k+j_k-n}$使得

$$P_{i_1+j_1-n} \subseteq \cdots \subseteq P_{i_k+j_k-n},$$

$$\dim P_{i_p+j_p-n} = i_p + j_p - n, \ p = 1, \cdots, k,$$

任取标准正交向量组$x_p \in P_{i_p+j_p-n}$, $p = 1, \cdots, k$, 则$x_p \in M_p, x_p \in N_p$. 由式(8.5.1)和式(8.5.2)得

$$\lambda_{i_1'}(A) + \cdots + \lambda_{i_k'}(A) \leqslant x_1^* A x_1 + \cdots x_k^* A x_k,$$

$$\lambda_{j_1'}(B) + \cdots + \lambda_{j_k'}(B) \leqslant x_1^* B x_1 + \cdots x_k^* B x_k,$$

于是

$$\begin{aligned} & x_1^*(A+B)x_1 + \cdots x_k^*(A+B)x_k \\ \geqslant \ & \lambda_{i_1'}(A) + \cdots + \lambda_{i_k'}(A) + \lambda_{j_1'}(B) + \cdots + \lambda_{j_k'}(B). \end{aligned}$$

使用$\{x_p\}$的任意性推出

$$\begin{aligned} & \inf_{\substack{\mu_p \in P_{i_p+j_p-n} \\ \{\mu_p\} o.n.}} (\mu_1^*(A+B)\mu_1 + \cdots + \mu_k^*(A+B)\mu_k) \\ \geqslant \ & \lambda_{i_1'}(A) + \cdots + \lambda_{i_k'}(A) + \lambda_{j_1'}(B) + \cdots + \lambda_{j_k'}(B). \end{aligned}$$

再由推论8.3.9得(i)成立.

(ii)类似于(i)的证明, 可由推论8.3.11证得结论. □

推论 8.5.1 设$A, B \in \mathbb{C}^{n\times n}$满足$A^* = A$且$B^* = B$, $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$ 是A的特征值, $\lambda_1(B) \geqslant \cdots \geqslant \lambda_n(B)$是$B$的特征值, $\lambda_1(A+B) \geqslant \cdots \geqslant \lambda_n(A+B)$是$A+B$的特征值, 正整数$j_1, \cdots, j_k$满足$1 \leqslant j_1 < \cdots < j_k \leqslant n$, 则

$$\lambda_n(A)+\cdots+\lambda_{n-k+1}(A) \leqslant \lambda_{j_1}(A+B)+\cdots+\lambda_{j_k}(A+B)-\lambda_{j_1}(B)-\cdots-\lambda_{j_k}(B), \tag{8.5.3}$$

$$\lambda_1(A)+\cdots+\lambda_k(A) \geqslant \lambda_{j_1}(A+B)+\cdots+\lambda_{j_k}(A+B)-\lambda_{j_1}(B)-\cdots-\lambda_{j_k}(B). \tag{8.5.4}$$

证明 在定理8.5.1(i)中取$i_p = n$, $p = 1, 2, \cdots, k$得式(8.5.3); 在定理8.5.1(ii)中取$i_p = 1$, $p = 1, 2, \cdots, k$得式(8.5.4). □

注记 8.5.1 若应用推论8.5.1及关系$-B = -(A+B) + A$可得出其他的不等式, 请读者自己写出.

定理 8.5.2 设$A, B \in \mathbb{C}^{n\times n}$, A有奇异值$\sigma_1(A) \geqslant \cdots \geqslant \sigma_n(A)$, B有奇异值$\sigma_1(B) \geqslant \cdots \geqslant \sigma_n(B)$, $A+B$有奇异值$\sigma_1(A+B) \geqslant \cdots \geqslant \sigma_n(A+B)$, 正整数$i_1, \cdots, i_k$满足$i_1 \leqslant \cdots \leqslant i_k$, 正整数$j_1, \cdots, j_k$满足$j_1 \leqslant \cdots \leqslant j_k$, 则

(i) 当$i_k \leqslant n$, $j_k \leqslant n$且$i_p + j_p \geqslant n + p$, $p = 1, 2, \cdots, k$时有

$$\begin{aligned} & \sigma_{i_1'}(A) + \cdots + \sigma_{i_k'}(A) + \sigma_{j_1'}(B) + \cdots + \sigma_{j_k'}(B) \\ \leqslant\ & \sigma_{(i_1+j_1-n)'}(A+B) + \cdots + \sigma_{(i_k+j_k-n)'}(A+B); \end{aligned}$$

(ii) 当$k \leqslant n$且$i_p + j_p \leqslant n - k + p + 1$, $p = 1, 2, \cdots, k$时有

$$\begin{aligned} & \sigma_{i_1''}(A) + \cdots + \sigma_{i_k''}(A) + \sigma_{j_1''}(B) + \cdots + \sigma_{j_k''}(B) \\ \geqslant\ & \sigma_{(i_1+j_1-1)''}(A+B) + \cdots + \sigma_{(i_k+j_k-1)''}(A+B). \end{aligned}$$

证明 结合引理8.1.1和定理8.5.1得到. □

习 题

1. 设$A^T = A \in \mathbb{R}^{n\times n}$, 证明:

$$\max_{\substack{y\in\mathbb{R}^n \\ \|y\|_2=1}} y^T A y = \max_{0\neq x\in\mathbb{R}^n} \frac{x^T A x}{x^T x} = \lambda_{\max}(A),$$

$$\min_{\substack{y\in\mathbb{R}^n\\ \|y\|_2=1}} y^T A y = \min_{0\neq x\in\mathbb{R}^n} \frac{x^T A x}{x^T x} = \lambda_{\min}(A).$$

2. 证明定理8.1.2(ii).

3. 证明定理8.1.3(ii).

4. 设A, $B\in\mathbb{C}^{n\times n}$, A有奇异值$\sigma_1(A)\geqslant\sigma_2(A)\geqslant\cdots\geqslant\sigma_n(A)$, B有奇异值$\sigma_1(B)\geqslant\sigma_2(B)\geqslant\cdots\geqslant\sigma_n(B)$, AB有奇异值$\sigma_1(AB)\geqslant\cdots\geqslant\sigma_n(AB)$, 若$i_1,\cdots,i_k$及$j_1,\cdots,j_k$为正整数, $k\leqslant n$, 证明:

   (i) 当$i_p+j_p\geqslant n+1$, $p=1,2,\cdots,k$时有
$$\begin{aligned}&\sigma_{i_1}(A)\cdots\sigma_{i_k}(A)\sigma_{j_1}(B)\cdots\sigma_{j_k}(B)\\ \leqslant\ &\sigma_{i_1+j_1-n}(AB)\cdots\sigma_{i_k+j_k-n}(AB);\end{aligned}$$

   (ii) 当$i_p+j_p\leqslant n+1$, $p=1,2,\cdots,k$时有
$$\begin{aligned}&\sigma_{i_1}(A)\cdots\sigma_{i_k}(A)\sigma_{j_1}(B)\cdots\sigma_{j_k}(B)\\ \geqslant\ &\sigma_{i_1+j_1-1}(AB)\cdots\sigma_{i_k+j_k-1}(AB).\end{aligned}$$

5. 设A, $B\in\mathbb{C}^{n\times n}$满足$A^*=A$且$B^*=B$, $\lambda_1(A)\geqslant\cdots\geqslant\lambda_n(A)$ 是A的特征值, $\lambda_1(B)\geqslant\cdots\geqslant\lambda_n(B)$是$B$的特征值, $\lambda_1(A+B)\geqslant\cdots\geqslant\lambda_n(A+B)$是$A+B$的特征值, 正整数$i_1,\cdots,i_k$满足$i_1\leqslant\cdots\leqslant i_k$, 正整数$j_1,\cdots,j_k$满足$j_1\leqslant\cdots\leqslant j_k$, 证明:

   (i) 当$i_k\leqslant n$, $j_k\leqslant n$且$i_p+j_p\geqslant n+1$, $p=1,2,\cdots,n$时有
$$\begin{aligned}&\lambda_{i_1}(A)+\cdots+\lambda_{i_k}(A)+\lambda_{j_1}(B)+\cdots+\lambda_{j_k}(B)\\ \leqslant\ &\lambda_{i_1+j_1-n}(A+B)+\cdots+\lambda_{i_k+j_k-n}(A+B);\end{aligned}$$

   (ii) 当$k\leqslant n$且$i_p+j_p\leqslant n+1$时有
$$\begin{aligned}&\lambda_{i_1}(A)+\cdots+\lambda_{i_k}(A)+\lambda_{j_1}(B)+\cdots+\lambda_{j_k}(B)\\ \geqslant\ &\lambda_{i_1+j_1-1}(A+B)+\cdots+\lambda_{i_k+j_k-1}(A+B).\end{aligned}$$

6. 设$A,B\in\mathbb{C}^{n\times n}$, $A=A^*$, $B=B^*$, 则$\rho(A+B)\leqslant\rho(A)+\rho(B)$.

7. 设$H^*=H\in\mathbb{C}^{n\times n}$的特征值为$\lambda_1\geqslant\cdots\geqslant\lambda_n$, v_i $(i=1,\cdots,n)$是H的属于λ_i的特征向量, 并且$v_1,\cdots,v_n$标准正交, $1\leqslant p\leqslant q\leqslant n$,
$$W=\operatorname{span}\{v_p\ ,v_{p+1}\ ,\cdots,\ v_q\},$$
   则对于任何的单位向量$x\in W$有$\lambda_q\leqslant x^*Hx\leqslant\lambda_p$.

8. 设$A\in\mathbb{C}^{n\times n}$, $\dfrac{A+A^*}{2}$的特征值为$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$, A的奇异值为$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n$, 则$\lambda_i\leqslant\sigma_i$, $i=1,\cdots,n$.

   (提示: 使用上题的结果, 设法寻找单位向量y使得$\lambda_k\leqslant y^*\dfrac{A+A^*}{2}y\leqslant\sqrt{y^*A^*Ay}\leqslant\sigma_k$.)

第9章　矩阵广义逆

考虑线性方程组

$$Ax = b, \tag{9.0.1}$$

其中$A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$. 若$m = n$且A可逆, 则线性方程组(9.0.1)对任意的$b \in \mathbb{C}^m$都有唯一解$x = A^{-1}b$; 否则, 线性方程组(9.0.1)不一定有解, 并且即使有解, 解也不一定唯一. 那么, 若线性方程组(9.0.1)有解, 是否可类似于A可逆时解的表达式$x = A^{-1}b$来表示方程组的解？当线性方程组(9.0.1)无解时, 是否可类似于A可逆时解的表达式$x = A^{-1}b$给出一个近似解？为了回答这两个问题, 本章介绍矩阵广义逆理论.

1920年E. H. Moore首次对$m \times n$矩阵给出了广义逆的定义: 对于矩阵$A \in \mathbb{C}^{m\times n}$, 若矩阵$X \in \mathbb{C}^{n\times m}$满足$AX = P_{\mathcal{R}(A)}$, $XA = P_{\mathcal{R}(X)}$, 则称X为A的Moore广义逆, 其中P_L是子空间L上的正交投影矩阵. 但是由于人们不知其用途, Moore广义逆一直未被重视. 直到20世纪50年代, R. Penrose以更明确的形式给出了它的另一个等价定义: $X \in \mathbb{C}^{n\times m}$是矩阵$A \in \mathbb{C}^{m\times n}$的Moore广义逆的充要条件是$AXA = A$, $XAX = X$, $(AX)^* = AX$, $(XA)^* = XA$. 此后, 人们开始称Moore广义逆为Moore-Penrose广义逆, 并且矩阵广义逆理论才逐步发展起来, 也广泛地应用于诸多学科, 如控制论、数理统计、多元分析、最优化理论等等.

应该强调的是, 可以根据实际需要来定义不同种类的矩阵广义逆, 如$\{i, j, \cdots, k\}$逆、Drazin逆、群逆、加权Drazin逆、加权Moore-Penrose逆等, 但所有矩阵广义逆都必须满足: 可逆阵的任一种广义逆均唯一且是其逆矩阵.

9.1　矩阵$\{i, j, \cdots, k\}$逆

9.1.1　矩阵$\{i, j, \cdots, k\}$逆的定义及其存在唯一性

定义 9.1.1　设$A \in \mathbb{C}^{m\times n}$, 若$X \in \mathbb{C}^{n\times m}$满足下面的方程(i)~(iv)的部分或全部, 称$X$为$A$的Penrose广义逆.

(i) $AXA=A$;

(ii) $XAX=X$;

(iii) $(AX)^*=AX$;

(iv) $(XA)^*=XA$.

由定义9.1.1知, A的Penrose广义逆共有$C_4^1+C_4^2+C_4^3+C_4^4=15$类. 称矩阵$A$的满足条件(i)~(iv)中$(i)$, (j), $\cdots$, (k)的广义逆为A的$\{i,j,\cdots,k\}$逆. 由于A的$\{i,j,\cdots,k\}$逆可能不唯一, 用$A\{i,j,\cdots,k\}$记A的所有$\{i,j,\cdots,k\}$逆构成的集合, 并用$A^{(i,j,\cdots,k)}$表示$A\{i,j,\cdots,k\}$中的任意确定的元素. 例如: $A\{1\}$表示A的所有$\{1\}$逆构成的集合, $A\{1,2,3,4\}$表示A的所有$\{1,2,3,4\}$逆构成的集合.

定理 9.1.1 设$A\in\mathbb{C}^{m\times n}$的奇异值分解为$A=U\mathrm{diag}(\Sigma,0)V$, 则

$$A\{(1)\}=\left\{V^*\begin{bmatrix}\Sigma^{-1} & X\\ Y & Z\end{bmatrix}U^*: X,Y,Z\text{任意}\right\},$$

$$A\{(1,2)\}=\left\{V^*\begin{bmatrix}\Sigma^{-1} & X\\ Y & Y\Sigma X\end{bmatrix}U^*: X,Y\text{任意}\right\},$$

$$A\{(1,3)\}=\left\{V^*\begin{bmatrix}\Sigma^{-1} & 0\\ Y & Z\end{bmatrix}U^*: Y,Z\text{任意}\right\},$$

$$A\{(1,4)\}=\left\{V^*\begin{bmatrix}\Sigma^{-1} & X\\ 0 & Z\end{bmatrix}U^*: X,Z\text{任意}\right\},$$

$$A\{(1,2,3)\}=\left\{V^*\begin{bmatrix}\Sigma^{-1} & 0\\ Y & 0\end{bmatrix}U^*: Y\text{任意}\right\},$$

$$A\{(1,2,4)\}=\left\{V^*\begin{bmatrix}\Sigma^{-1} & X\\ 0 & 0\end{bmatrix}U^*: X\text{任意}\right\},$$

$$A\{1,2,3,4\}=\left\{V^*\begin{bmatrix}\Sigma^{-1} & 0\\ 0 & 0\end{bmatrix}U^*\right\}.$$

证明 留给读者. □

上面的定理表明15类$\{i,j,\cdots,k\}$广义逆均存在, 下面证明$\{1,2,3,4\}$逆的唯一性(一般地, 其他的$\{i,j,\cdots,k\}$广义逆并不唯一).

命题 9.1.1 复矩阵的$\{1,2,3,4\}$逆唯一.

证明 设$X, Y \in A\{1,2,3,4\}$, 则

$$AXA = A, XAX = X, (AX)^* = AX, (XA)^* = XA,$$

$$AYA = A, YAY = Y, (AY)^* = AY, (YA)^* = YA.$$

于是

$$X = XAX = XAYAX = A^*X^*A^*Y^*X = A^*Y^*X = YAX,$$

同理$YAX = YAY$. 故$X = Y$, 即矩阵A的$\{1,2,3,4\}$逆唯一. □

为了简单起见, 下面将A的$\{1,2,3,4\}$逆记作A^+. 组合定理9.1.1和命题9.1.1有下面的推论.

推论 9.1.1 若$A \in \mathbb{C}^{m\times n}$有奇异值分解$A = U\mathrm{diag}(\Sigma, 0)V^*$, 则

$$A^+ = V\mathrm{diag}(\Sigma^{-1}, 0)U^*,$$

且A^+不依赖于V和U的选择.

定理 9.1.2 设$A \in \mathbb{C}^{m\times n}$, 则$X \in \mathbb{C}^{n\times m}$满足

$$AX = P_{\mathcal{R}(A)},\ XA = P_{\mathcal{R}(X)}$$

的充要条件是$X = A^+$.

证明 设$\mathrm{rank}A = r$, A有奇异值分解$A = U\mathrm{diag}(\Sigma, 0)V$,

$$X = V^*\begin{bmatrix} X_1 & X_2 \\ X_3 & X_4 \end{bmatrix}U^*.$$

充分性由推论9.1.1易见, 下证必要性. 易见

$$AX = U\begin{bmatrix} \Sigma X_1 & \Sigma X_2 \\ 0 & 0 \end{bmatrix}U^*, \tag{9.1.1a}$$

$$\mathcal{R}(A) = \left\{U\begin{bmatrix} y \\ 0 \end{bmatrix} : y \in \mathbb{C}^r\right\}, \mathcal{R}(A)^\perp = \left\{U\begin{bmatrix} 0 \\ z \end{bmatrix} : z \in \mathbb{C}^{n-r}\right\}. \tag{9.1.1b}$$

由$AX = P_{\mathcal{R}(A)}$得

$$AXv = P_{\mathcal{R}(A)}v = v,\quad \forall v \in \mathcal{R}(A),$$

$$AX\omega = P_{\mathcal{R}(A)}\omega = 0, \quad \forall \omega \in \mathcal{R}(A)^{\perp}.$$

由式(9.1.1)得$X_1 = \Sigma^{-1}, X_2 = 0$, 即

$$X = V^* \begin{bmatrix} \Sigma^{-1} & 0 \\ X_3 & X_4 \end{bmatrix} U^*. \tag{9.1.2}$$

故

$$XA = V^* \begin{bmatrix} I & 0 \\ X_3\Sigma & 0 \end{bmatrix} V, \tag{9.1.3a}$$

$$\mathcal{R}(X) = \left\{ V^* \begin{bmatrix} y_1 \\ X_3\Sigma y_1 + X_4 y_2 \end{bmatrix} : y_1 \in \mathbb{C}^r, y_2 \in \mathbb{C}^{m-r} \right\}. \tag{9.1.3b}$$

由$XA = P_{\mathcal{R}(X)}$得

$$XAv = P_{\mathcal{R}(X)}v = v, \quad \forall v \in \mathcal{R}(X), \tag{9.1.4}$$

$$XA\omega = P_{\mathcal{R}(X)}\omega = 0, \quad \forall \omega \in \mathcal{R}(X)^{\perp}. \tag{9.1.5}$$

组合式(9.1.3)和式(9.1.4)得$X_4 = 0$, 于是

$$X = V^* \begin{bmatrix} \Sigma^{-1} & 0 \\ X_3 & 0 \end{bmatrix} U^*, \mathcal{R}(X) = \left\{ V^* \begin{bmatrix} y_1 \\ X_3\Sigma y_1 \end{bmatrix} : y_1 \in \mathbb{C}^r \right\},$$

从而

$$\mathcal{R}(X)^{\perp} = \left\{ V^* \begin{bmatrix} -\Sigma X_3^* \beta \\ \beta \end{bmatrix} \middle| \beta \in \mathbb{C}^{m-r} \right\}.$$

这和式(9.1.3)及式(9.1.5)一起推出$X_3 = 0$, 故

$$X = V^* \mathrm{diag}(\Sigma^{-1}, 0) U^* = A^+.$$

□

基于定理9.1.2, 人们通常称A^+为A的Moore-Penrose广义逆. 上面定义了复矩阵的$\{i, j, \cdots, k\}$逆, 并解决了复矩阵的$\{i, j, \cdots, k\}$逆的存在唯一性问题. 真正地, 对于任意域(甚至环)上的矩阵, 仍可定义它的$\{1\}$逆、$\{2\}$逆和$\{1, 2\}$逆, 其定义方式与复数域上相同.

定理 9.1.3 设$A \in \mathbb{C}^{m\times n}$的等价分解为$A = P\mathrm{diag}(I_r, 0)Q$, 则

$$A\{1\} = \left\{ Q^{-1} \begin{bmatrix} I_r & X \\ Y & Z \end{bmatrix} P^{-1} : X, Y, Z\text{任意} \right\}, \tag{9.1.6}$$

$$A\{1, 2\} = \left\{ Q^{-1} \begin{bmatrix} I_r & X \\ Y & YX \end{bmatrix} P^{-1} : X, Y\text{任意} \right\}.$$

证明 留给读者. □

基于定理9.1.3, 矩阵A的$\{1\}$逆和$\{1,2\}$逆的表示可由下面的算法得到[2].

算法 9.1.1 (已知$A \in \mathbb{C}^{m\times n}$, 确定$A\{1\}$和$A\{1,2\}$)

(i) 构造矩阵$M = \begin{bmatrix} A & I_m \\ I_n & 0 \end{bmatrix}$;

(ii) 通过对矩阵M的前m行进行初等行变换和前n列进行初等列变换, 将M化成形式$\begin{bmatrix} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} & M_1 \\ M_2 & 0 \end{bmatrix}$;

(iii) 令

$$A\{1\} = \left\{ M_2 \begin{bmatrix} I & X \\ Y & Z \end{bmatrix} M_1 : X, Y, Z\text{任意} \right\},$$

$$A\{1,2\} = \left\{ M_2 \begin{bmatrix} I & X \\ Y & YX \end{bmatrix} M_1 : X, Y\text{任意} \right\}.$$

例 9.1.1 已知$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 0 & 1 \end{bmatrix}$, 求$A\{1\}$.

解

$$\begin{bmatrix} A & I_3 \\ I_2 & 0 \end{bmatrix} = \left[\begin{array}{cc|ccc} 1 & 2 & 1 & 0 & 0 \\ 2 & 4 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{cc|ccc} 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & -2 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}\right]$$

$$\to \left[\begin{array}{cc|ccc} 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ \hline 1 & -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{cc|ccc} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & -2 & 1 & 0 \\ \hline 1 & -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}\right],$$

故

$$M_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}, \ M_2 = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix},$$

因而

$$A\{1\} = \left\{ \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I_2 & X \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} : X \in \mathbb{C}^2 \right\}.$$

□

9.1.2 矩阵{1}逆和Moore-Penrose逆的性质

定理 9.1.4 设$A \in \mathbb{C}^{m\times n}$, 则

(i) $(A^*)\{1\} = \{X^* | X \in A\{1\}\}$;

(ii) $(\lambda A)\{1\} = \begin{cases} \mathbb{C}^{n\times m}, & 若\lambda = 0, \\ \lambda^{-1}A\{1\}, & 若\lambda \neq 0; \end{cases}$

(iii) $(PAQ)\{1\} = Q^{-1}A\{1\}P^{-1}, \forall P \in GL_m(\mathbb{C}), Q \in GL_n(\mathbb{C})$;

(iv) $\text{rank}A^{(1)} \geqslant \text{rank}A,\ \forall A^{(1)} \in A\{1\}$;

(v) $AA^{(1)}$和$A^{(1)}A$均为幂等阵, $\forall A^{(1)} \in A\{1\}$;

(vi) A列满秩$\Leftrightarrow A^{(1)}A = I_n,\ \forall A^{(1)} \in A\{1\}$;

(vii) A行满秩$\Leftrightarrow AA^{(1)} = I_m,\ \forall A^{(1)} \in A\{1\}$;

(viii) $A^{(1)}$唯一$\Leftrightarrow A$可逆;

(ix) $\text{rank}A = \text{rank}(AA^{(1)}) = \text{rank}(A^{(1)}A),\ \forall A^{(1)} \in A\{1\}$;

(x) $\mathcal{R}(A) = \mathcal{R}(AA^{(1)}), \mathcal{N}(A) = \mathcal{N}(A^{(1)}A),\ \forall A^{(1)} \in A\{1\}$.

证明 由矩阵{1}逆的定义和定理9.1.3易见. □

定理 9.1.5 设$A \in \mathbb{C}^{m\times n}$, 则

(i) $(A^+)^* = (A^*)^+, (A^+)^+ = A$;

(ii) $(A^*A)^+ = A^+(A^*)^+$;

(iii) $A^+ = (A^*A)^+A^*$;

(iv) $\text{rank}A = \text{rank}A^+$;

(v) 当U和V是酉矩阵时有$(UAV)^+ = V^*A^+U^*$.

证明 由推论9.1.1易见. □

定理 9.1.6 下面的命题成立:

(i) 若B列满秩, 则$B^+ = (B^*B)^{-1}B^*$;

(ii) 若C行满秩, 则$C^+ = C^*(CC^*)^{-1}$;

(iii) 若B列满秩且C行满秩, 则$(BC)^+ = C^+B^+$, 即

$$(BC)^+ = C^*(CC^*)^{-1}(B^*B)^{-1}B^*;$$

(iv) 若A的满秩分解为$A = BC$, 则$A^+ = C^+B^+$.

证明 (i) 由B列满秩知存在可逆阵P使得$B = P\begin{bmatrix} I_n \\ 0 \end{bmatrix}$, 于是

$$B^*B = \begin{bmatrix} I & 0 \end{bmatrix} PP^* \begin{bmatrix} I \\ 0 \end{bmatrix} > 0,$$

故B^*B可逆. 进而, 由矩阵Moore-Penrose广义逆的定义容易验证$B^+ = (B^*B)^{-1}B^*$.

(ii) 由C行满秩知C^*列满秩. 使用(i)推出$(C^*)^+ = (CC^*)^{-1}C$, 于是

$$C^+ = C^*(CC^*)^{-1}.$$

(iii) 由(i), (ii)及矩阵Moore-Penrose广义逆的定义易见.

(iv) 由(iii)直接得到. □

基于定理9.1.6(iii), 自然地有下面的问题:

问题 9.1.1 是否对任意矩阵$B \in \mathbb{C}^{m\times n}$和$C \in \mathbb{C}^{n\times p}$都有$(BC)^+ = C^+B^+$?

9.1.3 矩阵$\{1\}$逆的表示

定理9.1.1和定理9.1.3分别基于奇异值分解和等价分解给出了矩阵的几类$\{i,j,\cdots,k\}$逆的表示, 下面将展示如何用矩阵A的一个固定的$\{1\}$逆$A^{(1)}$来表示A的所有$\{1\}$逆.

定理 9.1.7 设$A\in\mathbb{C}^{m\times n}$, $A^{(1)}\in A\{1\}$, 则

$$A\{1\}=\left\{A^{(1)}+V-A^{(1)}AVAA^{(1)}:V\in\mathbb{C}^{n\times m}\right\}. \tag{9.1.7}$$

证明 任取$Y\in A\{1\}$, 则$AYA=A$. 取$V=Y-A^{(1)}$得

$$Y=A^{(1)}+V-A^{(1)}AVAA^{(1)}\in\left\{A^{(1)}+V-A^{(1)}AVAA^{(1)}:V\in\mathbb{C}^{n\times m}\right\},$$

于是$A\{1\}\subseteq\{A^{(1)}+V-A^{(1)}AVAA^{(1)}:V\in\mathbb{C}^{n\times m}\}$. 反之, 由

$$A(A^{(1)}+V-A^{(1)}AVAA^{(1)})A=A,\quad\forall V\in\mathbb{C}^{n\times m}$$

得$A\{1\}\supseteq\{A^{(1)}+V-A^{(1)}AVAA^{(1)}:V\in\mathbb{C}^{n\times m}\}$. □

定理 9.1.8 设$A\in\mathbb{C}^{m\times n}$, $A^{(1)}\in A\{1\}$, 则

$$A\{1\}=\left\{A^{(1)}+V(I_m-AA^{(1)})+(I_n-A^{(1)}A)W:V,W\in\mathbb{C}^{n\times m}\right\}. \tag{9.1.8}$$

证明 由

$$A\left[A^{(1)}+V(I_m-AA^{(1)})+(I_n-A^{(1)}A)W\right]A=A,\quad\forall V,W\in\mathbb{C}^{n\times m}$$

知$A\{1\}\supseteq\{A^{(1)}+V(I_m-AA^{(1)})+(I_n-A^{(1)}A)W:V,W\in\mathbb{C}^{n\times m}\}$. 反之, 对于任意的$Y\in A\{1\}$, 取$V=Y-A^{(1)}$且$W=-YAA^{(1)}$, 则

$$Y=A^{(1)}+V(I_m-AA^{(1)})-(I_n-A^{(1)}A)W,$$

故$A\{1\}\subseteq\{A^{(1)}+V(I_m-AA^{(1)})+(I_n-A^{(1)}A)W:V,W\in\mathbb{C}^{n\times m}\}$. □

9.1.4 矩阵$\{1\}$逆与矩阵方程的解

定理 9.1.9 设$A\in\mathbb{C}^{m\times n}$, $B\in\mathbb{C}^{m\times p}$, $A^{(1)}\in A\{1\}$, 则下列说法等价:

(i) $AX=B$有解.

(ii) $\mathcal{R}(B) \subseteq \mathcal{R}(A)$.

(iii) $AA^{(1)}B = B$.

(iv) $\operatorname{rank}\begin{bmatrix} A & B \end{bmatrix} = \operatorname{rank}A$.

证明

$$\begin{aligned}
AX = B\text{有解} \Leftrightarrow\ & B\text{的各列可由}A\text{的各列线性表出} \\
\Leftrightarrow\ & \mathcal{R}(B) \subseteq \mathcal{R}(A) \\
\Leftrightarrow\ & \text{存在}X_0\text{满足}B = AX_0 = AA^{(1)}AX_0 = AA^{(1)}B \\
\Leftrightarrow\ & \mathcal{R}(B) + \mathcal{R}(A) = \mathcal{R}(A) \\
\Leftrightarrow\ & \mathcal{R}\left(\begin{bmatrix} A & B \end{bmatrix}\right) = \mathcal{R}(A) \\
\Leftrightarrow\ & \operatorname{rank}\begin{bmatrix} A & B \end{bmatrix} = \operatorname{rank}A.
\end{aligned}$$

□

定理 9.1.10 设$A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{m\times p}$, $AX = B$有解, $A^{(1)} \in A\{1\}$, 则

(i) $A^{(1)}B$是$AX = B$的一个解;

(ii) $AX = 0$的通解表达式为$X = (I_n - A^{(1)}A)Y$, $\forall Y \in \mathbb{C}^{n\times p}$;

(iii) $AX = B$的通解表达式为$X = A^{(1)}B + (I_n - A^{(1)}A)Y$, $\forall Y \in \mathbb{C}^{n\times p}$.

证明 (i) 既然$AX = B$有解, 由定理9.1.9易见.

(ii) 由

$$A[(I_n - A^{(1)}A)Y] = (A - AA^{(1)}A)Y = 0, \quad \forall Y \in \mathbb{C}^{n\times p}$$

得所有形如$(I_n - A^{(1)}A)Y$的矩阵都是$AX = 0$的解. 另一方面, 任取$AX = 0$的解X_0, 则$AX_0 = 0$, 于是$X_0 = X_0 - A^{(1)}AX_0 = (I_n - A^{(1)}A)X_0$.

总之, $AX = 0$的通解为$X = (I_n - A^{(1)}A)Y$, $\forall Y \in \mathbb{C}^{n\times p}$.

(iii) 由(i)和(ii)知

$$A[A^{(1)}B + (I_n - A^{(1)}A)Y] = B, \forall Y \in \mathbb{C}^{n\times p},$$

即所有形如$A^{(1)}B + (I_n - A^{(1)}A)Y$的矩阵都是$AX = B$的解. 另一方面, 任取$AX = B$的解$X_1$, 则$AX_1 = B$. 使用(i)推出

$$A(X_1 - A^{(1)}B) = B - B = 0,$$

即$X_1 - A^{(1)}B$是$AX = 0$的解. 由(ii)知存在Y_1使得$X_1 - A^{(1)}B = (I_n - A^{(1)}A)Y_1$, 故

$$X_1 = A^{(1)}B + (I_n - A^{(1)}A)Y_1.$$

总之, $AX=B$的通解表达式为$X=A^{(1)}B+(I_n-A^{(1)}A)Y,\ \forall Y\in\mathbb{C}^{n\times p}$. □

类似于上面的两个定理容易得到下面的两个定理.

定理 9.1.11 设$A\in\mathbb{C}^{m\times n}$, $C\in\mathbb{C}^{q\times n}$, $A^{(1)}\in A\{1\}$, 则下列说法等价:

(i) $XA=C$有解;

(ii) $CA^{(1)}A=C$;

(iii) $\mathcal{R}(C^*)\subseteq\mathcal{R}(A^*)$ (即$\mathcal{N}(C)\supseteq\mathcal{N}(A)$);

(iv) $\operatorname{rank}\begin{bmatrix}A\\C\end{bmatrix}=\operatorname{rank}A$.

定理 9.1.12 设$A\in\mathbb{C}^{m\times n}$, $C\in\mathbb{C}^{q\times n}$, $XA=C$有解, $A^{(1)}\in A\{1\}$, 则

(i) $CA^{(1)}$是$XA=C$的解;

(ii) $XA=0$的通解表达式为$X=Z(I_m-AA^{(1)}),\ \ \forall Z\in\mathbb{C}^{q\times m}$;

(iii) $XA=C$的通解表达式为$X=CA^{(1)}+Z(I_m-AA^{(1)}),\ \ \forall Z\in\mathbb{C}^{q\times m}$.

定理 9.1.13 设$A\in\mathbb{C}^{m\times n},B\in\mathbb{C}^{m\times p}$且$AX=B$有解. 若$B$列满秩, 则$AX=B$的通解为$X=GB,\forall G\in A\{1\}$.

证明 由定理9.1.10(i)知对任意的$G\in A\{1\}$, GB是$AX=B$的解.

任取A的$\{1\}$逆$A^{(1)}$和$AX=B$的解X_0, 由定理9.1.10(iii)得存在$Y_0\in\mathbb{C}^{n\times p}$使得

$$X_0=A^{(1)}B+(I_n-A^{(1)}A)Y_0. \tag{9.1.9}$$

由B列满秩知$\operatorname{rank}\begin{bmatrix}B\\Y_0\end{bmatrix}=\operatorname{rank}B$. 使用定理9.1.11推出$Y_0=ZB$有解, 即存在$Z_0\in\mathbb{C}^{n\times m}$使得$Y_0=Z_0B$. 因此式(9.1.9)变成

$$X_0=A^{(1)}B+(I_n-A^{(1)}A)Z_0B=G_0B,$$

其中$G_0=A^{(1)}+(I_n-A^{(1)}A)Z_0$. 由

$$AG_0A=A[A^{(1)}+(I_n-A^{(1)}A)Z_0]A=AA^{(1)}A+A(I_n-A^{(1)}A)Z_0A=A$$

得$G_0\in A\{1\}$. 总之, $AX=B$的通解为$X=GB,\forall G\in A\{1\}$. □

9.1.5 矩阵$\{1,4\}$逆与线性方程组的最小范数解

考虑线性方程组

$$Ax = b, \tag{9.1.10}$$

其中$A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$.

定义 9.1.2 设线性方程组(9.1.10)有解, 称所有解中2−范数最小者(即满足$\sqrt{x^*x}$最小的解x) 为方程组(9.1.10)的最小范数解.

定理 9.1.14 设方程组(9.1.10)有解, $A^{(1,4)} \in A\{1,4\}$, 则$A^{(1,4)}b$为方程组(9.1.10)的最小范数解.

证明 设矩阵A的奇异值分解为

$$A = U\text{diag}(\Sigma, 0)V,$$

由方程组(9.1.10)有解及定理9.1.9得

$$b = U\begin{bmatrix} b_1 \\ 0 \end{bmatrix}, \tag{9.1.11}$$

并且线性方程组(9.1.10)的通解表示为

$$x = V^*\begin{bmatrix} \Sigma^{-1}b_1 \\ w \end{bmatrix}, \ \forall w.$$

再由定理9.1.1推出

$$A^{(1,4)}b = V^*\begin{bmatrix} \Sigma^{-1}b_1 \\ 0 \end{bmatrix}$$

是线性方程组(9.1.10)的最小范数解. □

注意到$V^*\begin{bmatrix} \Sigma^{-1}b_1 \\ 0 \end{bmatrix} = A^+b$, 从上面定理的证明容易得到下面的命题.

命题 9.1.2 若线性方程组(9.1.10)有解, 则A^+b是它的唯一最小范数解.

9.1.6 矩阵$\{1,3\}$逆与线性方程组的最小二乘解

定义 9.1.3 若$x_0 \in \mathbb{C}^n$满足

$$\|b - Ax_0\|_2 = \min_{x\in\mathbb{C}^n} \|b - Ax\|_2,$$

称x_0是线性方程组(9.1.10)的最小二乘解.

从上面的定义易见, 若线性方程组(9.1.10)有解, 则它的每个解都是最小二乘解.

定理 9.1.15 设$A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$, $A^{(1,3)} \in A\{1,3\}$, 则$A^{(1,3)}b$是线性方程组(9.1.10)的最小二乘解.

证明 设矩阵A的奇异值分解为$A = U\mathrm{diag}(\Sigma, 0)V$, 并设

$$b = U\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad x = V^*\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

于是

$$b - Ax = U\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} - U\mathrm{diag}(\Sigma, 0)Vx = U\begin{bmatrix} b_1 - \Sigma x_1 \\ b_2 \end{bmatrix},$$

因而

$$\min_{x\in\mathbb{C}^n} \|b - Ax\|_2 = \|b_2\|_2. \tag{9.1.12}$$

另一方面, 由定理9.1.1得

$$\begin{aligned} & \|b - AA^{(1,3)}b\|_2 \\ = & \left\| U\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} - U\mathrm{diag}(\Sigma, 0)VV^*\begin{bmatrix} \Sigma^{-1} & 0 \\ Y & Z \end{bmatrix} U^*U\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right\|_2 \\ = & \left\| U\begin{bmatrix} 0 \\ b_2 \end{bmatrix} \right\|_2 \\ = & \|b_2\|_2. \end{aligned}$$

这和式(9.1.12)一起推出$A^{(1,3)}b$是线性方程组(9.1.10)的最小二乘解. □

从上面定理的证明容易看出, 线性方程组(9.1.10)的所有最小二乘解可表示为:

$$x = V^*\begin{bmatrix} \Sigma^{-1}b_1 \\ w \end{bmatrix}, \quad \forall w. \tag{9.1.13}$$

由此容易推出下面的命题.

命题 9.1.3 x是线性方程组(9.1.10)的最小二乘解的充要条件是x是线性方程组$A^*Ax = A^*b$的解.

9.1.7 矩阵M-P逆与线性方程组的最小范数最小二乘解

定义 9.1.4 称线性方程组(9.1.10)的所有最小二乘解中2−范数最小者为最小范数最小二乘解.

组合命题9.1.2和式(9.1.13)易得下面的定理.

定理 9.1.16 线性方程组(9.1.10)有唯一的最小范数最小二乘解A^+b.

9.1.8 Schur补与分块矩阵的{1}–逆

对于矩阵

$$M=\begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

当A可逆时称$D-CA^{-1}B$为M关于A的Schur补, 记作M/A; 当D可逆时称$A-BD^{-1}C$为M关于D的Schur补, 记作M/D.

命题 9.1.4 [3] 设$M=\begin{bmatrix} A & B \\ C & D \end{bmatrix}$可逆且$A$可逆, 则

$$M^{-1}=\begin{bmatrix} X & Y \\ U & V \end{bmatrix},$$

其中

$$\begin{aligned} X &= A^{-1}+A^{-1}B\,(M/A)^{-1}\,CA^{-1}, \\ Y &= -A^{-1}B\,(M/A)^{-1}, \\ U &= -\,(M/A)^{-1}\,CA^{-1}, \\ V &= (M/A)^{-1}. \end{aligned}$$

命题 9.1.5 设$A\in\mathbb{C}^{n\times n}, A>0$, 则

$$(A_{i_1\cdots i_k})^{-1}\leqslant (A^{-1})_{i_1\cdots i_k},$$

其中$A_{i_1\cdots i_k}$表示A的子阵

$$\begin{bmatrix} a_{i_1i_1} & a_{i_1i_2} & \cdots & a_{i_1i_k} \\ a_{i_2i_1} & a_{i_2i_2} & \cdots & a_{i_2i_k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_ki_1} & a_{i_ki_2} & \cdots & a_{i_ki_k} \end{bmatrix}.$$

证明 不妨设

$$A=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},\quad A_{11}=A_{i_1\cdots i_k}.$$

由命题9.1.4得

$$(A^{-1})_{i_1\cdots i_k}=A_{11}^{-1}+A_{11}^{-1}A_{12}(A/A_{11})^{-1}A_{21}A_{11}^{-1}\geqslant A_{11}^{-1}=(A_{i_1\cdots i_k})^{-1}.$$

□

例 9.1.2 已知$A=\begin{bmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{bmatrix}$, 其中$A_{11}$可逆, 求$A$的一个$\{1\}$–逆.

解 易见

$$A=\begin{bmatrix}I&0\\A_{21}A_{11}^{-1}&I\end{bmatrix}\begin{bmatrix}A_{11}&0\\0&A/A_{11}\end{bmatrix}\begin{bmatrix}I&A_{11}^{-1}A_{12}\\0&I\end{bmatrix}.$$

由定理9.1.4(iii)得

$$\begin{bmatrix}I&-A_{11}^{-1}A_{12}\\0&I\end{bmatrix}\begin{bmatrix}A_{11}&0\\0&(A/A_{11})^{(1)}\end{bmatrix}\begin{bmatrix}I&0\\-A_{21}A_{11}^{-1}&I\end{bmatrix}\in A\{1\},$$

于是

$$\begin{aligned}&\begin{bmatrix}I&-A_{11}^{-1}A_{12}\\0&I\end{bmatrix}\begin{bmatrix}A_{11}^{-1}&0\\0&(A/A_{11})^{(1)}\end{bmatrix}\begin{bmatrix}I&0\\-A_{21}A_{11}^{-1}&I\end{bmatrix}\\=&\begin{bmatrix}A_{11}^{-1}+A_{11}^{-1}A_{12}(A/A_{11})^{(1)}A_{21}A_{11}^{-1}&-A_{11}^{-1}A_{12}(A/A_{11})^{(1)}\\-(A/A_{11})^{(1)}A_{21}A_{11}^{-1}&(A/A_{11})^{(1)}\end{bmatrix}\in A\{1\}.\end{aligned}$$

□

上面构造$A^{(1)}$的过程的一个关键步骤是将A分解成形式$A=PDQ$, 其中P,Q可逆, D是对角块阵. A_{11}可逆只是为了保证这个分解能够实现. 那么, 当A_{11}不可逆, 需要加什么条件才能保证这个分解能够实现呢? 为此, 考察

$$M:=\begin{bmatrix}I&0\\X&I\end{bmatrix}\begin{bmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{bmatrix}\begin{bmatrix}I&Y\\0&I\end{bmatrix}$$

是对角块阵的条件? 由

$$M=\begin{bmatrix}A_{11}&A_{12}+A_{11}Y\\A_{21}+XA_{11}&A_{22}+XA_{12}+(A_{21}+XA_{11})Y\end{bmatrix}$$

知

$$M\text{是对角块阵}\Leftrightarrow X\text{是}XA_{11}=-A_{21}\text{的解且}Y\text{是}A_{11}Y=-A_{12}\text{的解}.$$

使用定理9.1.9和定理9.1.11推出: 当

$$A_{21}A_{11}^{(1)}A_{11}=A_{21},\quad A_{11}A_{11}^{(1)}A_{12}=A_{12}\tag{9.1.14}$$

时, 若取$(X,Y)=(-A_{21}A_{11}^{(1)},-A_{11}^{(1)}A_{12})$, 则

$$M=\mathrm{diag}(A_{11},\widehat{A/A_{11}}),$$

其中$\widehat{A/A_{11}}=A_{22}-A_{21}A_{11}^{(1)}A_{12}$.

称$\widehat{A/A_{11}}$为A的关于A_{11}的广义Schur补. 因而, 类似于例9.1.2有:

命题 9.1.6 当式(9.1.14)成立时,

$$\begin{bmatrix} A_{11}^{(1)}+A_{11}^{(1)}A_{12}(\widehat{A/A_{11}})^{(1)}A_{21}A_{11}^{(1)} & -A_{11}^{(1)}A_{12}(\widehat{A/A_{11}})^{(1)} \\ -(\widehat{A/A_{11}})^{(1)}A_{21}A_{11}^{(1)} & (\widehat{A/A_{11}})^{(1)} \end{bmatrix}\in A\{1\}.$$

例 9.1.3 已知$A=\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{bmatrix}\geqslant 0$, 其中$A_{11}$为方阵, 求$A$的一个$\{1\}$–逆.

解 由$A\geqslant 0$知, 存在行满秩矩阵B使得$A=B^*B$. 令$B=\begin{bmatrix} B_1 & B_2 \end{bmatrix}$, 则

$$A=\begin{bmatrix} B_1^* \\ B_2^* \end{bmatrix}\begin{bmatrix} B_1 & B_2 \end{bmatrix}=\begin{bmatrix} B_1^*B_1 & B_1^*B_2 \\ B_2^*B_1 & B_2^*B_2 \end{bmatrix},$$

于是$A_{11}=B_1^*B_1$, $A_{12}=B_1^*B_2$, $A_{22}=B_2^*B_2$, 从而

$$\begin{aligned} A_{12}^*A_{11}^{(1)}A_{11} &= (B_1^*B_2)^*(B_1^*B_1)^{(1)}(B_1^*B_1) \\ &= B_2^*B_1(B_1^*B_1)^{(1)}B_1^*B_1 \\ &= B_2^*B_1 \\ &= A_{12}^* \end{aligned}$$

且

$$A_{11}A_{11}^{(1)}A_{12}=B_1^*B_1(B_1^*B_1)^{(1)}B_1^*B_2=B_1^*B_2=A_{12}.$$

使用命题9.1.6推出

$$\begin{bmatrix} A_{11}^{(1)}+A_{11}^{(1)}A_{12}(\widehat{A/A_{11}})^{(1)}A_{12}^*A_{11}^{(1)} & -A_{11}^{(1)}A_{12}(\widehat{A/A_{11}})^{(1)} \\ -(\widehat{A/A_{11}})^{(1)}A_{12}^*A_{11}^{(1)} & (\widehat{A/A_{11}})^{(1)} \end{bmatrix}\in A\{1\}.$$

□

9.2 矩阵Drazin逆

9.2.1 矩阵Drazin逆的定义及其存在唯一性

定义 9.2.1 设$A\in\mathbb{F}^{n\times n}$, 称满足$\mathrm{rank}A^{k+1}=\mathrm{rank}A^k$的最小非负整数$k$为$A$的指标, 记作$\mathrm{Ind}(A)$.

定义 9.2.2 设$A\in\mathbb{F}^{n\times n}$, $k=\mathrm{Ind}(A)$, 若$X\in\mathbb{F}^{n\times n}$满足

$$A^{k+1}X=A^k,\ XAX=X,\ AX=XA,$$

称X是A的一个Drazin逆, 简称$D-$逆.

下面的定理给出了矩阵Drazin逆的存在性与唯一性.

定理 9.2.1 设$A\in\mathbb{F}^{n\times n}$, 则$A$的Drazin逆存在且唯一.

证明 若A可逆, 易见A^{-1}是A的唯一的Drazin逆, 即A的Drazin逆存在且唯一. 若A不可逆, 设A的Fitting分解为

$$A=P\mathrm{diag}(D,N)P^{-1},$$

则容易验证$P\mathrm{diag}(D^{-1},0)P^{-1}$是$A$的Drazin逆, 并且在$A$的Fitting分解给定的情况下, A的Drazin逆是确定的. 因而, 为了展示A的Drazin逆是唯一的, 只需证: 对于A的任意两个Fitting分解

$$A=P_1\mathrm{diag}(D_1,N_1)P_1^{-1}=P_2\mathrm{diag}(D_2,N_2)P_2^{-1}$$

均有

$$P_1\mathrm{diag}(D_1^{-1},0)P_1^{-1}=P_2\mathrm{diag}(D_2^{-1},0)P_2^{-1}.$$

事实上, 由$P_1\mathrm{diag}(D_1,N_1)P_1^{-1}=P_2\mathrm{diag}(D_2,N_2)P_2^{-1}$得

$$P_2^{-1}P_1\mathrm{diag}(D_1,N_1)=\mathrm{diag}(D_2,N_2)P_2^{-1}P_1.$$

令

$$P_2^{-1}P_1=\begin{bmatrix} M_1 & M_2\\ M_3 & M_4\end{bmatrix},$$

则$M_1D_1=D_1M_1$, $M_3D_1=N_2M_3$且$M_2N_1=D_2M_2$, 于是

$$M_1D_1^{-1}=D_1^{-1}M_1,\ M_3=0,\ M_2=0.$$

故$P_1\mathrm{diag}(D_1^{-1},0)P_1^{-1}=P_2\mathrm{diag}(D_2^{-1},0)P_2^{-1}$. □

将矩阵A的Drazin逆记作A_d. 从定理9.2.1的证明可知下面的命题成立.

命题 9.2.1 设$A\in\mathbb{F}^{n\times n}$的Fitting分解为$A=P\mathrm{diag}(D,N)P^{-1}$, 则

$$A_d=P\mathrm{diag}(D^{-1},0)P^{-1}.$$

推论 9.2.1 设$A\in\mathbb{F}^{n\times n}$, 若$A$可逆, 则$A_d=A^{-1}$; 若$A$幂零, 则$A_d=0$.

9.2.2 矩阵Drazin逆的性质

在1.2中，我们介绍了矩阵Fitting分解的定义，它可用来证明关于矩阵Drazin逆的很多结论，下面是几个例子.

命题 9.2.2 设$A \in \mathbb{F}^{n\times n}$, $\mathrm{Ind}(A) = k$, 则存在多项式$q(x)$使得$A_d = A^k[q(A)]^{k+1}$.

证明 设A的Fitting分解为$A = P\mathrm{diag}(D, N)P^{-1}$, 则$N^k = 0$(为什么?). 取

$$f(x) = \det(xI - D) := x^s + a_1x^{s-1} + \cdots + a_{s-1}x + a_s,$$

由D可逆及Hamilton-Cayley定理知$a_s \neq 0$且$f(D) = 0$. 由此可推出

$$D^{-1} = -\frac{1}{a_s}(D^{s-1} + a_1D^{s-2} + \cdots + a^{s-1}I),$$

即存在一个多项式$q(x)$使得$D^{-1} = q(D)$, 于是

$$\begin{aligned} A^k[q(A)]^{k+1} &= P\mathrm{diag}(D^k, 0)P^{-1}P\mathrm{diag}(q^{k+1}(D), q^{k+1}(N))P^{-1} \\ &= P\mathrm{diag}(D^kD^{-(k+1)}, 0)P^{-1} \\ &= P\mathrm{diag}(D^{-1}, 0)P^{-1} \\ &= A_d. \end{aligned}$$

□

命题 9.2.3 设$A \in \mathbb{F}^{n\times n}$, $A = \mathrm{diag}(A_1, \cdots, A_s)$, A_1, $\cdots$, A_s为方阵, 则$A_d = \mathrm{diag}((A_1)_d, \cdots, (A_s)_d)$.

证明 设A_i的Fitting分解为

$$A_i = P_i\mathrm{diag}(D_i, N_i)P_i^{-1}, \ i = 1, 2, \cdots, s,$$

则

$$\begin{aligned} A &= P_1\mathrm{diag}(D_1, N_1)P_1^{-1}, \cdots, P_s\mathrm{diag}(D_s, N_s)P_s^{-1} \\ &= P\mathrm{diag}(D_1, N_1, \cdots, D_s, N_s)P^{-1}, \end{aligned}$$

其中$P = \mathrm{diag}(P_1, \cdots, P_s)$. 于是存在置换阵$Q$使得

$$A = PQ\mathrm{diag}(D_1, D_2, \cdots, D_s, N_1, N_2, \cdots, N_s)(PQ)^{-1},$$

从而

$$\begin{aligned} A_d = &\ PQ\mathrm{diag}(D_1^{-1}, D_2^{-1}, \cdots, D_s^{-1}, 0)(PQ)^{-1} \\ = &\ P\mathrm{diag}(D_1^{-1}, 0, D_2^{-1}, 0, \cdots, D_s^{-1}, 0)P^{-1} \\ = &\ \mathrm{diag}(P_1\mathrm{diag}((D_1^{-1}, 0)P_1^{-1}, \cdots, P_s\mathrm{diag}(D_s^{-1}, 0)P_s^{-1}) \\ = &\ \mathrm{diag}((A_1)_d, (A_2)_d, \cdots, (A_s)_d). \end{aligned}$$

□

命题 9.2.4 设$A \in \mathbb{F}^{n\times n}$, 则$(A_d)_d = A$的充要条件是$\mathrm{Ind}(A) = 1$.

证明 **必要性**. 设A的Fitting分解为

$$A = P\mathrm{diag}(D, N)P^{-1},$$

由命题9.2.1得

$$(A_d)_d = P\mathrm{diag}(D, 0)P^{-1}.$$

因为$(A_d)_d = A$, 所以$N = 0$, 由此可得$\mathrm{Ind}(A) = 1$.

充分性. 由$\mathrm{Ind}(A) = 1$得A的Fitting分解为$A = P\mathrm{diag}(D, 0)P^{-1}$, 显然有$(A_d)_d = A$. □

命题 9.2.5 若$A, B \in \mathbb{F}^{n\times n}$满足$AB = BA$, 则$(AB)_d = B_d A_d$.

证明 (i) 当A, B均可逆时显然成立.

(ii) 当A不可逆且B可逆时. 设$\mathrm{Ind}(A) = h$且A的Fitting分解为$A = P\mathrm{diag}(D, N)P^{-1}$. 由定理9.2.1的证明可设

$$B = P\mathrm{diag}(B_1, B_4)P^{-1},$$

其中B_1和B_4满足$DB_1 = B_1 D$且

$$NB_4 = B_4 N. \tag{9.2.1}$$

故

$$AB = P\mathrm{diag}(D, N)P^{-1}P\mathrm{diag}(B_1, B_4)P^{-1} = P\mathrm{diag}(DB_1, NB_4)P^{-1}.$$

因为B可逆, 所以B_1, B_4可逆, 从而$(DB_1)_d = B_1^{-1}D^{-1}$. 又因为$N$为幂零阵, 所以由式(9.2.1)推出$NB_4$为幂零阵. 因而

$$\begin{aligned}(AB)_d = & \ P\mathrm{diag}(B_1^{-1}D^{-1}, 0)P^{-1}\\ = & \ P\mathrm{diag}(B_1^{-1}, B_4^{-1})P^{-1}P\mathrm{diag}(D^{-1}, 0)P^{-1}.\\ = & \ B^{-1}A_d\\ = & \ B_d A_d.\end{aligned}$$

(iii) 当A可逆且B不可逆时, 类似于(ii)可以证明结论成立.

(iv) 当A, B均不可逆时, 设A的Fitting分解为$A = P\mathrm{diag}(D, N)P^{-1}$. 由定理9.2.1的证明可设

$$B = P\mathrm{diag}(B_1, B_4)P^{-1},$$

其中B_1和B_4满足$DB_1 = B_1 D$且

$$NB_4 = B_4 N, \tag{9.2.2}$$

于是

$$(AB)_d = P\text{diag}((DB_1)_d, (NB_4)_d)P^{-1}, \quad B_dA_d = P\text{diag}\left((B_1)_dD^{-1}, 0\right)P^{-1}.$$

故只需证$(DB_1)_d = (B_1)_dD^{-1}$且NB_4为幂零阵. 事实上, 由N为幂零阵及式(9.2.2)易见NB_4为幂零阵. 若B_1可逆, 则由(i)和D可逆得

$$(DB_1)_d = B_1^{-1}D^{-1} = (B_1)_dD^{-1};$$

若B_1不可逆, 则由(iii)和D可逆得$(DB_1)_d = (B_1)_dD^{-1}$. 总之, $(AB)_d = B_dA_d$. □

9.2.3 矩阵群逆

本节介绍一种特殊的Drazin逆—群逆.

定义 9.2.3 设$A \in \mathbb{F}^{n\times n}$, 如果$X \in \mathbb{F}^{n\times n}$满足

$$AXA = A,\ XAX = X,\ AX = XA,$$

称X为A的群逆.

由定义可以看出: 当$\text{Ind}(A) = 1$时, A的群逆就是Drazin逆. 由上节可知, 对任意的$A \in \mathbb{F}^{n\times n}$, A_d存在且唯一, 但是群逆可能是不存在的. 例如:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

定理 9.2.2 设$A \in \mathbb{F}^{n\times n}$且$A$为奇异阵, 则

(i) A有群逆的充要条件是$\text{Ind}(A) = 1$ (即存在非奇异阵P和D使得$A = P\text{diag}(D, 0)P^{-1}$);

(ii) 当群逆存在时, 它是唯一的.

证明 类似于定理9.2.1(留给读者). □

当矩阵A的群逆存在时, 将其记作A_g.

推论 9.2.2 设$A \in \mathbb{F}^{n\times n}$, A_g存在, 则存在可逆阵$P \in \mathbb{F}^{n\times n}$和$D \in \mathbb{F}^{r\times r}$使得$A = P\text{diag}(D, O)P^{-1}$且$A_g = P\text{diag}(D^{-1}, O)P^{-1}$.

证明 由定理9.2.2和命题9.2.1得到. □

既然群逆是Drazin逆的特例, 由命题9.2.2和推论9.2.2易得下面的命题.

命题 9.2.6 设$A \in \mathbb{F}^{n\times n}$满足$\text{Ind}(A) = 1$, 则

(i) 存在一个多项式$q(x)$使得$A_g = A[q(x)]^2$;

(ii) $\mathcal{R}(A_g) = \mathcal{R}(A)$;

(iii) $\mathcal{N}(A_g) = \mathcal{N}(A)$.

由命题9.2.5和定理9.2.2容易推出下面的结论.

命题 9.2.7 设$A, B \in \mathbb{F}^{n\times n}$, A_g和B_g存在, $AB = BA$, 则$(AB)_g$存在且

$$(AB)_g = B_g A_g. \tag{9.2.3}$$

称式(9.2.3)为群逆的逆序律. 文献[15]给出了逆序律成立的充要条件.

习 题

1. 设$A \in \mathbb{R}^{m\times n}$, 证明: $P_{\mathcal{R}(A)} = A(A^TA)^{(1)}A^T = P^T_{\mathcal{R}(A)}$, $P^2_{\mathcal{R}(A)} = P_{\mathcal{R}(A)}$.

2. 设$H \in \mathbb{C}^{n\times n}$, 证明: $H^+ = H$的充要条件是H^2为幂等Hermite矩阵且
$$\text{rank}H^2 = \text{rank}H.$$
(提示: 使用H的Fitting分解或使用Moore-Penrose广义逆的定义)

3. 设矩阵$A = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix}$, 求$A\{1\}$.

4. 设$A \in \mathbb{F}^{m\times n}$, $B \in \mathbb{F}^{n\times p}$, 则

   (i) $\text{rank}(AB) = \text{rank}A$的充要条件是$ABXA = A$, $\forall X \in (AB)\{1\}$;

   (ii) $\text{rank}(AB) = \text{rank}B$的充要条件是$BXAB = B$, $\forall X \in (AB)\{1\}$.

   (提示: 使用定理3.4.1)

5. 设$A \in \mathbb{C}^{m\times n}$, 则

   (i) $A(A^*A)^{(1)}A^*$与$(A^*A)^{(1)}$的选取无关;

(ii) $A(A^*A)^{(1)}A^*A = A, A^*A(A^*A)^{(1)}A^* = A^*$.

6. 设$A \in \mathbb{C}^{n\times n}$, $B = [A^T \quad A^T]^T$, 试证: $B^+ = \frac{1}{2}[A^+ \quad A^+]$.

7. 设$A_i \in \mathbb{C}^{m\times n}, b_i \in \mathbb{C}^m \ (i = 1, 2, \cdots, k)$, 证明: 向量$x_0 \in \mathbb{C}^n$使得

$$\sum_{i=1}^{k} \|A_i x_0 - b_i\|_2^2 = \min_{x\in\mathbb{C}^n} \sum_{i=1}^{k} \|A_i x - b_i\|_2^2$$

的充要条件是x_0为方程

$$\left(\sum_{i=1}^{k} A_i^* A_i\right) x = \sum_{i=1}^{k} A_i^* b_i$$

的解.

8. 已知$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$.

(i) 求A的奇异值分解;

(ii) 利用奇异值分解求A^+;

(iii) 求$Ax = b$的最小范数解.

9. 给定方程组$Ax = b$, 其中

$$A = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix}, b = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix},$$

求它的最小范数最小二乘解.

10. 设$A \in \mathbb{F}^{n\times n}$,

$$\lambda^+ = \begin{cases} \lambda^{-1}, & \lambda \neq 0, \\ 0, & \lambda = 0. \end{cases}$$

证明: λ是A的特征值的充要条件是λ^+是A_d的特征值.

11. 证明: 对任意的$A \in \mathbb{F}^{n\times n}$有$(A^T)_d = (A_d)^T$.

12. 证明: 对任意的$A \in \mathbb{F}^{n\times n}$有$[(A_d)_d]_d = A_d$.

13. 设$A, B \in \mathbb{F}^{n\times n}$满足$A = PBP^{-1}$, 则$A_d = PB_dP^{-1}$.

14. 设$A \in \mathbb{F}^{n\times n}$, s是任意非负整数, 则$(A^s)_d = (A_d)^s$.

15. 设$A \in \mathbb{F}^{n\times n}$, 若整数$k$和$m$满足$k > m > 0$, 则$A^m(A_d)^k = (A_d)^{k-m}$.

16. 设$A \in \mathbb{F}^{n\times n}$, $\mathrm{Ind}(A) = k$, 则

$$A_d = A^l X A^l,\ \forall l \geqslant k,\ X \in A^{2l+1}\{1\}.$$

17. 设$A \in \mathbb{F}^{n\times n}$满足$\mathrm{Ind}(A) = 1$, 证明:

    (i) $AA_g = A_gA = P_{\mathcal{R}(A_g),\mathcal{N}(A_g)} = P_{\mathcal{R}(A),\mathcal{N}(A)}$;

    (ii) $I - AA_g = I - A_gA = P_{\mathcal{N}(A),\mathcal{R}(A)}$;

    (iii) 若$A = BC$是A的一个满秩分解, 则$A_g = B(CB)^{-2}C$.

18. 设$A \in \mathbb{F}^{n\times n}$, 证明: 若$A$非奇异, 则$\mathrm{Ind}(A) = 0$; 若$A$奇异, 则$\mathrm{Ind}(A) \geqslant 1$.

19. 设$N \in \mathbb{F}^{n\times n}$幂零, 称满足$N^k = 0$的最小正整数$k$为$N$的幂零指标. 证明: 若奇异矩阵$A \in \mathbb{F}^{n\times n}$的Fitting分解为$A = P\mathrm{diag}(D, N)P^{-1}$, 则$\mathrm{Ind}(A)$就是$N$的幂零指标.

20. 设$A, B \in \mathbb{F}^{n\times n}$满足$A$奇异, $AB = BA$, $ABA = A$, 则$\mathrm{Ind}(A) = 1$.

第10章　矩阵的Kronecker积和Hadamard积

本章主要介绍矩阵的Kronecker积和Hadamard积的定义和主要性质，以及矩阵的Kronecker积在线性矩阵方程求解中的应用.

10.1　矩阵的Kronecker积的定义和性质

以前在线性代数中定义过两个矩阵A与B的乘积，它要求A的列数与B的行数必须相等. 下面我们引进一种新的矩阵乘法，它对矩阵A与B的行数与列数没有任何要求.

定义 10.1.1　设$A=[a_{ij}]\in\mathbb{C}^{m\times n}$, $B\in\mathbb{C}^{p\times q}$, 称$mp\times nq$矩阵

$$\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$$

为矩阵A与B的Kronecker积，记作$A\otimes B$.

例 10.1.1　设$A=\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B=\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$, 则

$$A\otimes B=\begin{bmatrix} 0 & 1 & 0 & 2 \\ 2 & 3 & 4 & 6 \\ 0 & 3 & 0 & 4 \\ 6 & 9 & 8 & 12 \end{bmatrix},\ B\otimes A=\begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 3 & 4 \\ 2 & 4 & 3 & 6 \\ 6 & 8 & 9 & 12 \end{bmatrix}.$$

上例表明运算$\otimes$不满足交换律, 即$A\otimes B\neq B\otimes A$.

性质 10.1.1 设$A, A_1, A_2, B, B_1, B_2, C$和$D$是有合适维数的复矩阵, 则

(i) $(A_1+A_2)\otimes B=A_1\otimes B+A_2\otimes B,\ A\otimes(B_1+B_2)=A\otimes B_1+A\otimes B_2$;

(ii) $(aA)\otimes B=a(A\otimes B)=A\otimes(aB),\ \ \forall\, a\in\mathbb{C}$;

(iii) $\overline{A\otimes B}=\overline{A}\otimes\overline{B},\ \ (A\otimes B)^T=A^T\otimes B^T,\ \ (A\otimes B)^*=A^*\otimes B^*$;

(iv) $0\otimes B=A\otimes 0=0$;

(v) $A\otimes B=0\Rightarrow A=0$或$B=0$;

(vi) $(A\otimes B)(C\otimes D)=AC\otimes BD$.

证明 通过直接计算得到. □

命题 10.1.1 设$A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{m\times m}$, 若$A$, B可逆, 则$A\otimes B$也可逆, 且$(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$.

证明 由性质10.1.1(vi)得

$$(A\otimes B)\left(A^{-1}\otimes B^{-1}\right)=AA^{-1}\otimes BB^{-1}=I_n\otimes I_m=I_{mn},$$

故$A\otimes B$可逆, 且$(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$. □

命题 10.1.2 若$\mathrm{rank}A=r$, $\mathrm{rank}B=s$, 则$\mathrm{rank}(A\otimes B)=rs$.

证明 设

$$A=P_1\mathrm{diag}(I_r,0)Q_1, B=P_2\mathrm{diag}(I_s,0)Q_2,$$

其中P_i, $Q_i(i=1,2)$为可逆阵. 由性质10.1.1(vi)得

$$\begin{aligned}A\otimes B= &\ (P_1\mathrm{diag}(I_r,0)Q_1)\otimes(P_2\mathrm{diag}(I_s,0)Q_2)\\ = &\ (P_1\otimes P_2)\left(\mathrm{diag}(I_r,0)\otimes\mathrm{diag}(I_s,0)\right)(Q_1\otimes Q_2).\end{aligned}$$

这和定义10.1.1及命题10.1.1一起推出$\mathrm{rank}(A\otimes B)=rs$. □

命题 10.1.3 设$A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{m\times m}$, 则$\mathrm{tr}(A\otimes B)=(\mathrm{tr}A)(\mathrm{tr}B)$.

证明 由Jordan标准形理论可设

$$A=P\begin{bmatrix} a_1 & * & \cdots & \cdots & * \\ 0 & a_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & * \\ 0 & \cdots & \cdots & 0 & a_n \end{bmatrix}P^{-1},$$

$$B=Q\begin{bmatrix} b_1 & * & \cdots & \cdots & * \\ 0 & b_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & * \\ 0 & \cdots & \cdots & 0 & b_m \end{bmatrix}Q^{-1},$$

则

$$A\otimes B=(P\otimes Q)\begin{bmatrix} a_1b_1 & * & \cdots & \cdots & \cdots & * \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & a_1b_m & \ddots & & \vdots \\ \vdots & & \ddots & a_2b_1 & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & * \\ 0 & \cdots & \cdots & \cdots & 0 & a_nb_m \end{bmatrix}(P\otimes Q)^{-1},$$

所以$\mathrm{tr}(A\otimes B)=\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}a_ib_j=\left(\sum\limits_{i=1}^{n}a_i\right)\left(\sum\limits_{j=1}^{m}b_j\right)=(\mathrm{tr}A)(\mathrm{tr}B)$. □

类似于命题10.1.3的证明易推出下面的定理.

定理 10.1.1 设$A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{m\times m}$, A的特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$, B的特征值为$\mu_1,\mu_2,\cdots,\mu_m$, 则$A\otimes B$的特征值为$\lambda_i\mu_j$, $i=1,2,\cdots,n$, $j=1,2,\cdots,m$.

推论 10.1.1 设$A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{m\times m}$, A的特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$, B的特征值为$\mu_1,\mu_2,\cdots,\mu_m$, 则$A^k\otimes B^l$的特征值为

$$\lambda_i^k\mu_j^l,\ i=1,2,\cdots,n,\quad j=1,2,\cdots,m.$$

推论 10.1.2 设$A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{m\times m}$, 则

$$\det(A\otimes B)=(\det A)^m(\det B)^n.$$

推论 10.1.3 若$A \geqslant 0$, $B \geqslant 0$, 则$A \otimes B \geqslant 0$.

定理 10.1.2 设$A_1, A_2 \in \mathbb{C}^{m\times n}$, $B_1, B_2 \in \mathbb{C}^{s\times t}$为同型阵. 若$A_1 \otimes B_1 = A_2 \otimes B_2 \neq 0$, 则存在非零常数$\lambda$使得$A_1 = \lambda A_2$, $B_1 = \lambda^{-1} B_2$.

证明 记$A_1 = [a_{ij}^{(1)}]_{m\times n}$, $A_2 = [a_{ij}^{(2)}]_{m\times n}$, 则

$$A_1 \otimes B_1 = \begin{bmatrix} a_{11}^{(1)}B_1 & a_{12}^{(1)}B_1 & \cdots & a_{1n}^{(1)}B_1 \\ a_{21}^{(1)}B_1 & a_{22}^{(1)}B_1 & \cdots & a_{2n}^{(1)}B_1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}^{(1)}B_1 & a_{m2}^{(1)}B_1 & \cdots & a_{mn}^{(1)}B_1 \end{bmatrix},$$

$$A_2 \otimes B_2 = \begin{bmatrix} a_{11}^{(2)}B_2 & a_{12}^{(2)}B_2 & \cdots & a_{1n}^{(2)}B_2 \\ a_{21}^{(2)}B_2 & a_{22}^{(2)}B_2 & \cdots & a_{2n}^{(2)}B_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}^{(2)}B_2 & a_{m2}^{(2)}B_2 & \cdots & a_{mn}^{(2)}B_2 \end{bmatrix}.$$

由$A_1 \otimes B_1 = A_2 \otimes B_2 \neq 0$知存在$p, q$使得$a_{pq}^{(1)}B_1 = a_{pq}^{(2)}B_2 \neq 0$. 令$\lambda = \dfrac{a_{pq}^{(1)}}{a_{pq}^{(2)}}$, 则$\lambda \neq 0$且$B_2 = \lambda B_1$, 于是由$A_1 \otimes B_1 = A_2 \otimes B_2$得$A_1 \otimes B_1 = \lambda(A_2 \otimes B_1)$, 从而$(A_1 - \lambda A_2) \otimes B_1 = 0$. 由性质10.1.1(v)得$A_1 = \lambda A_2$. □

定理 10.1.3 $A \otimes B$是非零的正规矩阵的充要条件是A和B均为非零正规矩阵.

证明 **充分性**. 由A和B均为正规阵知$AA^* = A^*A$, $BB^* = B^*B$, 于是

$$\begin{aligned}(A\otimes B)(A\otimes B)^* &= (A\otimes B)(A^*\otimes B^*) \\ &= AA^*\otimes BB^* \\ &= A^*A\otimes B^*B \\ &= (A^*\otimes B^*)(A\otimes B) \\ &= (A\otimes B)^*(A\otimes B),\end{aligned}$$

故$A \otimes B$是正规矩阵. 再由$A, B \neq 0$知$A \otimes B \neq 0$.

必要性. 设

$$A = U\begin{bmatrix} a_1 & * & \cdots & * \\ 0 & a_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix}U^*, B = V\begin{bmatrix} b_1 & * & \cdots & * \\ 0 & b_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_m \end{bmatrix}V^*,$$

其中U, V为酉矩阵, 则

$$A \otimes B = (U \otimes V)C(U \otimes V)^*,$$

其中

$$C = \begin{bmatrix} a_1b_1 & * & \cdots & \cdots & \cdots & * \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & a_1b_m & \ddots & & \vdots \\ \vdots & & \ddots & a_2b_1 & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & * \\ 0 & \cdots & \cdots & \cdots & 0 & a_nb_m \end{bmatrix}.$$

由$A \otimes B$是正规矩阵知C为正规矩阵, 于是$CC^* = C^*C$, 所以C为对角阵, 从而A, B相似于对角阵. 再由$A \otimes B \neq 0$知A和B均为非零正规矩阵. □

命题 10.1.4　若$A \otimes B$是Hermite正定阵, 则

$$A = \varepsilon A_0,\ B = \varepsilon B_0,$$

其中$\varepsilon = \pm 1$, A_0, B_0是Hermite正定阵.

证明　由$A \otimes B$是Hermite正定阵知$A \otimes B$是可逆正规矩阵, 从而由定理10.1.3和命题10.1.1知A和B均为可逆正规矩阵. 设

$$A = U\mathrm{diag}(a_1, a_2, \cdots, a_n)U^*, B = V\mathrm{diag}(b_1, b_2, \cdots, b_m)V^*,$$

其中$a_i, b_i \neq 0$, $i = 1, 2, \cdots, n$, U, V是酉阵, 则

$$A \otimes B = (U \otimes V)\mathrm{diag}(a_1b_1, \cdots, a_1b_m, a_2b_1, \cdots, a_nb_m)(U \otimes V)^*. \quad (10.1.1)$$

由$A \otimes B$是Hermite正定阵知$a_ib_j > 0$, $\forall i, j$, 从而$a_1b_j > 0$, $\forall j$, 因此$b_1, \cdots, b_m$同号. 故$a_1, \cdots, a_n, b_1, \cdots, b_m$同号. 再由$A \otimes B$正定及式(10.1.1)知, 存在$\varepsilon \in \{1, -1\}$使得$A = \varepsilon A_0, B = \varepsilon B_0$, 其中

$$A_0 = U\mathrm{diag}(|a_1|, |a_2|, \cdots, |a_n|)U^*, B_0 = V\mathrm{diag}(|b_1|, |b_2|, \cdots, |b_m|)V^*.$$

□

定义 10.1.2　设$P \in \mathbb{R}^{n \times n}$, 若$P$中的每行每列只有一个1, 且$PP^T = I_n$, 称$P$为置换阵.

命题 10.1.5　存在置换阵P使得

$$A \otimes I_n = P(I_n \otimes A)P^T, \quad \forall A \in \mathbb{C}^{m \times m},$$

$$I_m \otimes B = P(B \otimes I_m)P^T, \quad \forall B \in \mathbb{C}^{n \times n}.$$

证明 由Kronecker积的定义及数学归纳法易见. □

性质 10.1.2 设$A \in \mathbb{C}^{m\times m}$, $B \in \mathbb{C}^{n\times n}$, 则$A\otimes B$置换相似于$B\otimes A$.

证明 由命题10.1.5知存在置换阵P使得

$$\begin{aligned} A\otimes B = & (A\otimes I_n)(I_m\otimes B) \\ = & P(I_n\otimes A)P^TP(B\otimes I_m)P^T \\ = & P(I_n\otimes A)(B\otimes I_m)P^T \\ = & P(B\otimes A)P^T, \end{aligned}$$

即$A\otimes B$置换相似于$B\otimes A$. □

10.2 矩阵的Kronecker积与线性矩阵方程的解

前面已经定义了矩阵空间上的拉直映射. 矩阵的Kronecker积有下面的重要性质:

性质 10.2.1 设$A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{n\times p}$, $C \in \mathbb{C}^{p\times q}$, 则

$$\mathrm{Vec}(ABC) = \left(C^T\otimes A\right)\mathrm{Vec}B.$$

证明 记$B = [b_1\ b_2\ \cdots\ b_p]$, $C = [c_1\ c_2\ \cdots\ c_q] = [c_{ij}]$, 则

$$\begin{aligned} \mathrm{Vec}(ABC) = & \begin{bmatrix} ABc_1 \\ ABc_2 \\ \vdots \\ ABc_q \end{bmatrix} = \begin{bmatrix} A[b_1\ b_2\ \cdots\ b_p]c_1 \\ A[b_1\ b_2\ \cdots\ b_p]c_2 \\ \vdots \\ A[b_1\ b_2\ \cdots\ b_p]c_q \end{bmatrix} \\ = & \begin{bmatrix} A(c_{11}b_1+\cdots+c_{p1}b_p) \\ A(c_{12}b_1+\cdots+c_{p2}b_p) \\ \vdots \\ A(c_{1q}b_1+\cdots+c_{pq}b_p) \end{bmatrix} \\ = & \begin{bmatrix} c_{11}A & c_{21}A & \cdots & c_{p1}A \\ c_{12}A & c_{22}A & \cdots & c_{p2}A \\ \vdots & \vdots & \ddots & \vdots \\ c_{1q}A & c_{2q}A & \cdots & c_{pq}A \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = \left(C^T\otimes A\right)\mathrm{Vec}B. \end{aligned}$$

□

定理 10.2.1 矩阵方程$A_1XB_1+A_2XB_2+\cdots+A_nXB_n=C$有解的充要条件为

$$\begin{aligned}&\operatorname{rank}[B_1^T\otimes A_1+B_2^T\otimes A_2+\cdots+B_n^T\otimes A_n]\\=\ &\operatorname{rank}[B_1^T\otimes A_1+B_2^T\otimes A_2+\cdots+B_n^T\otimes A_n\quad \operatorname{Vec}C].\end{aligned}$$

证明 由性质10.2.1知

$$\begin{aligned}&\text{矩阵方程}A_1XB_1+A_2XB_2+\cdots+A_nXB_n=C\text{有解}\\\Leftrightarrow\ &\operatorname{Vec}(A_1XB_1)+\operatorname{Vec}(A_2XB_2)+\cdots+\operatorname{Vec}(A_nXB_n)=\operatorname{Vec}C\text{有解}\\\Leftrightarrow\ &[(B_1^T\otimes A_1)+(B_2^T\otimes A_2)+\cdots+(B_n^T\otimes A_n)]\operatorname{Vec}X=\operatorname{Vec}C\text{有解}\\\Leftrightarrow\ &\operatorname{rank}[B_1^T\otimes A_1+B_2^T\otimes A_2+\cdots+B_n^T\otimes A_n]\\&=\operatorname{rank}[B_1^T\otimes A_1+B_2^T\otimes A_2+\cdots+B_n^T\otimes A_n\quad \operatorname{Vec}C].\end{aligned}$$

□

推论 10.2.1 矩阵方程$AXB=D$有解的充要条件为

$$\operatorname{rank}\left(B^T\otimes A\right)=\operatorname{rank}[B^T\otimes A\quad \operatorname{Vec}D].$$

推论 10.2.2 若$B^T\otimes A$可逆, 则矩阵方程$AXB=D$有唯一解

$$X=\operatorname{Vec}^{-1}\left(\left(B^T\otimes A\right)^{-1}\operatorname{Vec}D\right).$$

推论 10.2.3 矩阵方程$AX+XB=D$有解的充要条件为

$$\operatorname{rank}\left(I\otimes A+B^T\otimes I\right)=\operatorname{rank}\left[\begin{array}{cc}I\otimes A+B^T\otimes I & \operatorname{Vec}D\end{array}\right].$$

定理 10.2.2 设$A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{m\times m}$, A的特征根为$\lambda_1,\lambda_2,\cdots,\lambda_n$, B的特征根为$\mu_1,\mu_2,\cdots,\mu_m$, 则$\lambda_i+\mu_j$是$\left(I_m\otimes A+B^T\otimes I_n\right)$的特征根, $i=1,2,\cdots,n,j=1,2,\cdots,m$.

证明 设$x_1,x_2,\cdots,x_n$分别为A的特征根$\lambda_1,\lambda_2,\cdots,\lambda_n$对应的特征向量, $y_1,y_2,\cdots,y_m$分别为B^T的特征根$\mu_1,\mu_2,\cdots,\mu_m$对应的特征向量, 则

$$Ax_i=\lambda_ix_i, i=1,2,\cdots,n,$$

$$B^Ty_j=\mu_jy_j,\ j=1,2,\cdots,m.$$

要证$\lambda_i+\mu_j$是$I\otimes A+B^T\otimes I$的特征根, 只需证存在非零向量z_{ij}使得

$$\left(I\otimes A+B^T\otimes I\right)z_{ij}=(\lambda_i+\mu_j)z_{ij}.$$

事实上, 对任意的i和j, 令$z_{ij} = y_j \otimes x_i$, 则$z \neq 0$且

$$\begin{aligned}\left(I \otimes A + B^T \otimes I\right) z_{ij} &= (I \otimes A)(y_j \otimes x_i) + (B^T \otimes I)(y_j \otimes x_i) \\ &= (y_j \otimes Ax_i) + (B^T y_j \otimes x_i) \\ &= \lambda_i(y_j \otimes x_i) + \mu_j(y_j \otimes x_i) \\ &= (\lambda_i + \mu_j)(y_j \otimes x_i).\end{aligned}$$

□

推论 10.2.4 设$A \in \mathbb{C}^{n\times n}$的特征根为$\lambda_1$, λ_2, $\cdots$, λ_n, $B \in \mathbb{C}^{m\times m}$的特征根为$\mu_1$, μ_2, $\cdots$, μ_m, 则$I \otimes A + B^T \otimes I$可逆的充分必要条件是

$$\lambda_i + \mu_j \neq 0,\ i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m.$$

组合推论10.2.3和推论10.2.4易见:

定理 10.2.3 若A的任意两个特征根的和不是0, 则矩阵方程$AX + XA^T = D$的解存在且唯一.

10.3 矩阵的Hadamard积

定义 10.3.1 设$A, B \in \mathbb{C}^{m\times n}$且$A = [a_{ij}]$, $B = [b_{ij}]$, 称$m \times n$矩阵

$$\begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2n}b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \cdots & a_{mn}b_{mn} \end{bmatrix}$$

为矩阵A与B的Hadamard积, 记作$A \circ B$.

由矩阵的Hadamard积的定义显然有

$$A \circ B = B \circ A,\quad (A + B) \circ C = (A \circ C) + (B \circ C),$$

并且

$$A \circ B = P(A \otimes B)Q, \tag{10.3.1}$$

其中

$$P = E_{11} + E_{2,m+2} + \cdots + E_{m,m^2} \in \mathbb{R}^{m\times m^2},$$

$$Q = E_{11} + E_{n+2,2} + \cdots + E_{n^2,n} \in \mathbb{R}^{n^2\times n}.$$

特别地, 若$m = n$, 则$Q = P^T$, 且

$$A \circ B = P(A \otimes B)P^T. \tag{10.3.2}$$

因而$A \circ B$是$A \otimes B$的主子阵. 故有下面的命题.

命题 10.3.1 设$A, B \in \mathbb{C}^{m\times n}$, $\text{rank}A = r_1$, $\text{rank}B = r_2$, 则$\text{rank}(A \circ B) \leqslant r_1 r_2$.

命题 10.3.2 设$A, B \in \mathbb{C}^{n\times n}$, $A, B \geqslant 0$, 则

$$\lambda_{\min}(A \circ B) \geqslant \lambda_{\min}(A)\lambda_{\min}(B).$$

证明 设$A \otimes B = Q \begin{bmatrix} A \circ B & * \\ * & * \end{bmatrix} Q^T$, 其中$Q$为置换阵, 由Sturm分离原理及定理10.1.1知结论成立. □

命题 10.3.3 设$A, B \in \mathbb{C}^{n\times n}$, 若$A > 0, B > 0$, 则$A \circ B > 0$.

证明 由推论10.1.3及$A \circ B$为$A \otimes B$的主子阵得证. □

命题 10.3.4 设$A, B \in \mathbb{C}^{n\times n}$,

$$D = \text{diag}(d_1, d_2, \cdots, d_n), \ E = \text{diag}(e_1, e_2, \cdots, e_n),$$

则$(DA) \circ (BE) = A \circ (DBE) = (DAE) \circ B$.

证明 由矩阵的Hadamard积的定义得

$$(DA) \circ (BE) = D(A \circ B)E = (A \circ (DB))E = A \circ (DBE) = (DAE) \circ B.$$

□

定义 10.3.2 若n阶矩阵H的每个元素均为1或-1, 且$H^T H = nI_n$, 称H是Hadamard矩阵.

易见矩阵$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$和$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$均为Hadamard矩阵.

命题 10.3.5 设H_1和H_2是Hadamard矩阵, 则$H_1 \otimes H_2$是Hadamard矩阵.

证明 由Kronecker积的性质及Hadamard矩阵的定义得

$$\begin{aligned} (H_1 \otimes H_2)^T (H_1 \otimes H_2) = & \left(H_1^T \otimes H_2^T\right)(H_1 \otimes H_2) \\ = & \left(H_1^T H_1\right) \otimes \left(H_2^T H_2\right) \\ = & mI_m \otimes nI_n \\ = & mnI_{mn}, \end{aligned}$$

其中m和n分别是H_1和H_2的阶数. □

命题 10.3.6 不存在3阶Hadamard矩阵.

命题 10.3.7 设$A, B \in \mathbb{C}^{n\times n}$, $A > 0$, $B > 0$, 证明:

(i) $(A\circ B)^{-1} \leqslant A^{-1}\circ B^{-1}$;

(ii) $\det(A\circ B) \geqslant (\det A)(\det B)$.

证明 (i) 由式(10.3.2)和命题9.1.5得

$$\begin{aligned}(A\circ B)^{-1} \leqslant\ & P(A\otimes B)^{-1}P^T\\ =\ & P(A^{-1}\otimes B^{-1})P^T\\ =\ & A^{-1}\circ B^{-1}.\end{aligned}$$

(ii) 对n使用数学归纳法. 当$n=1$时, 显然; 假设$n=k$时结论成立. 当$n=k+1$时, 令

$$A = \begin{bmatrix} A_1 & \alpha \\ \alpha^T & a \end{bmatrix}, B = \begin{bmatrix} B_1 & \beta \\ \beta^T & b \end{bmatrix},$$

则

$$(\det A)(\det B) = (\det A_1)(\det B_1)(a-\alpha^T A_1^{-1}\alpha)(b-\beta^T B_1^{-1}\beta). \quad (10.3.3)$$

进而由(i)得

$$\begin{aligned}&\det(A\circ B)\\ =\ & \det(A_1\circ B_1)\left(ab-(\alpha^T\circ\beta^T)(A_1\circ B_1)^{-1}(\alpha\circ\beta)\right)\\ \geqslant\ & \det(A_1\circ B_1)\left(ab-(\alpha^T\circ\beta^T)(A_1^{-1}\circ B_1^{-1})(\alpha\circ\beta)\right)\\ =\ & \det(A_1\circ B_1)\left(ab-(\alpha^T\otimes\beta^T)P^TP(A_1^{-1}\otimes B_1^{-1})P^TP(\alpha\otimes\beta)\right).\end{aligned}$$

使用$P^TP = E_{11}+E_{m+2,m+2}+\cdots+E_{m^2,m^2}$推出

$$\begin{aligned}\det(A\circ B) \geqslant\ & \det(A_1\circ B_1)\left(ab-(\alpha^T\otimes\beta^T)(A_1^{-1}\otimes B_1^{-1})(\alpha\otimes\beta)\right)\\ =\ & \det(A_1\circ B_1)\left(ab-(\alpha^T A_1^{-1}\alpha)(\beta^T B_1^{-1}\beta)\right).\end{aligned}$$

由式(10.3.3)和归纳假设, 只需证

$$2(\alpha^T A_1^{-1}\alpha)(\beta^T B_1^{-1}\beta) < a\beta^T B_1^{-1}\beta + b\alpha^T A_1^{-1}\alpha,$$

亦只需证$2 \leqslant \dfrac{a}{\alpha^T A_1^{-1}\alpha} + \dfrac{b}{\beta^T B_1^{-1}\beta}$, 只需证$\alpha^T A_1^{-1}\alpha < a$且$\beta^T B_1^{-1}\beta < b$. 事实上, 这由$A > 0$及$B > 0$容易推出. □

习　题

1. 若A, B为实对称阵, 则$A\otimes B$也为实对称阵.

2. 若A, B为Hermite阵, 则$A\otimes B$为Hermite阵.

3. 设A和B均为幂等阵, 则$A\otimes B$为幂等阵.

4. 证明: A, B为对角阵的充要条件是$0\neq A\otimes B$为对角阵.

5. 证明: A, B为上三角阵的充要条件是$0\neq A\otimes B$为上三角阵.

6. 设$A\in\mathbb{C}^{m\times n}$, $B\in\mathbb{C}^{p\times q}$, 则$(A\otimes B)^+=A^+\otimes B^+$.

7. 设$A\in\mathbb{C}^{m\times n}$, $B\in\mathbb{C}^{p\times q}$, k是正整数, 证明: $(AB)^{[k]}=A^{[k]}B^{[k]}$, 其中

$$A^{[k]}=\underbrace{A\otimes A\cdots\otimes A}_{k}.$$

(提示: 对k应用数学归纳法.)

8. 设$X\in\mathbb{R}^{2\times 2}$,

$$A=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix},\ B=\begin{bmatrix}1 & 0 & -1\\ 2 & 1 & 0\end{bmatrix},\ D=\begin{bmatrix}0 & 1 & -1\\ -1 & 0 & 1\end{bmatrix}.$$

判断矩阵方程$AXB=D$是否有解?

9. 证明: 矩阵方程$AXB=D$有唯一解的充要条件为

$$\text{rank}\left(B^T\otimes A\right)=\text{rank}[B^T\otimes A\ \ \text{Vec}D]=B^T\otimes A\text{的列数}.$$

10. 设

$$A=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix},\ B=\begin{bmatrix}3 & -2\\ 1 & 0\end{bmatrix},\ D=\begin{bmatrix}-2 & 4\\ -3 & 3\end{bmatrix},$$

证明: 矩阵方程$AX-XB=D$有唯一解, 并求它的解.

11. 设

$$A=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\ B=\begin{bmatrix}1 & 2 & 1\\ 0 & -1 & 2\\ 0 & 0 & -1\end{bmatrix},\ D=\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\end{bmatrix},$$

证明: 矩阵方程$X-AXB=D$有唯一解, 并求它的解.

12. 设$A,D\in\mathbb{C}^{n\times n}$, $\rho(A)<1$, 则矩阵方程$X-AXA^*=D$有唯一解$X=\sum\limits_{i=0}^{+\infty}A^iD(A^*)^i$.

13. 证明: 矩阵方程组

$$\begin{cases} A_1X_1B_1 + A_2X_2B_2 + \cdots + A_nX_nB_n = C, \\ C_1X_1D_1 + C_2X_2D_2 + \cdots + C_nX_nD_n = E. \end{cases}$$

有解的充要条件为

$$\begin{aligned} &\operatorname{rank}\begin{bmatrix} B_1^T \otimes A_1 & B_2^T \otimes A_2 & \cdots & B_n^T \otimes A_n \\ D_1^T \otimes C_1 & D_2^T \otimes C_2 & \cdots & D_n^T \otimes C_n \end{bmatrix} \\ = \ &\operatorname{rank}\begin{bmatrix} B_1^T \otimes A_1 & B_2^T \otimes A_2 & \cdots & B_n^T \otimes A_n & \mathrm{Vec}C \\ D_1^T \otimes C_1 & D_2^T \otimes C_2 & \cdots & D_n^T \otimes C_n & \mathrm{Vec}E \end{bmatrix}. \end{aligned}$$

14. 设$X \in \mathbb{C}^m$, $Y \in \mathbb{C}^n$, $\|X\|_2 = \|Y\|_2 = 1$, 求$\|X \otimes Y\|_2$.

第11章 线性矩阵不等式

一个线性矩阵不等式就是具有形式

$$F(x) := F_0 + x_1F_1 + \cdots + x_mF_m < 0 \tag{11.0.1}$$

的一个表达式, 其中$x_1, \cdots, x_m$是实变量(称为线性矩阵不等式(11.0.1)的决策变量), $F_i = F_i^T \in \mathbb{R}^{n\times n}, i = 0, 1, \cdots, m$是一组给定的实对称矩阵, 小于号"<"指的是矩阵$F(x)$是负定的, 即对所有非零的向量$v \in \mathbb{R}^n$有$v^TF(x)v < 0$. 称

$$x = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}^T \in \mathbb{R}^m$$

为决策向量(即由决策变量构成的向量).

在许多系统与控制问题中, 决策变量是以矩阵的形式出现的. 例如Lyapunov矩阵不等式:

$$A^TX + XA + Q < 0, \tag{11.0.2}$$

其中$A, Q \in \mathbb{R}^{n\times n}$是给定的常数矩阵, 且$Q$是对称的, $X \in \mathbb{R}^{n\times n}$是对称的未知矩阵变量. 设$E_1, E_2, \cdots, E_m$是所有实$n$阶对称矩阵空间的基, 则存在$x_1, x_2, \cdots, x_m$使得

$$X = \sum_{i=1}^{m} x_iE_i.$$

因此Lyapunov矩阵不等式(11.0.2)可写成线性矩阵不等式的形式

$$Q + x_1(A^TE_1 + E_1A) + \cdots + x_m(A^TE_m + E_mA) < 0.$$

对于给定的实对称矩阵$F_0, F_1, \cdots, F_m$, 容易证明集合

$$\{x : F(x) < 0\}$$

是一个凸集. 因此形如线性矩阵不等式的约束条件是对变量的一个凸约束. 这一性质使得可以应用解决凸优化问题的有效方法来求解线性矩阵不等式, 例如: 椭圆方法, 内点方法等. 值得提到的是: 越来越多的系统与控制中的问题的求解被转化为求解线性矩阵不等式, 并且可以使用MATLAB的LMI工具箱来求解线性矩阵不等式.

本章介绍一些重要的线性矩阵不等式.

11.1 Schur补引理及其应用

Schur补引理在使用线性矩阵不等式方法处理许多问题中都有应用，本节介绍Schur补引理以及几个应用的例子.

引理 11.1.1 (Schur补引理) 对给定的实对称矩阵$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{bmatrix}$，其中$S_{11}$是方阵，则以下三个命题是等价的:

(i) $S < 0$;

(ii) $S_{11} < 0,\ S_{22} - S_{12}^T S_{11}^{-1} S_{12} < 0$;

(iii) $S_{22} < 0,\ S_{11} - S_{12} S_{22}^{-1} S_{12}^T < 0$.

证明 (i)⇔(ii) 应用块矩阵的初等运算，可以得到

$$\begin{bmatrix} I & 0 \\ -S_{12}^T S_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ -S_{12}^T S_{11}^{-1} & I \end{bmatrix}^T = \begin{bmatrix} S_{11} & 0 \\ 0 & S_{22} - S_{12}^T S_{11}^{-1} S_{12} \end{bmatrix}.$$

因此

$$(i)\text{成立} \Leftrightarrow \begin{bmatrix} S_{11} 0 & \\ 0 & S_{22} - S_{12}^T S_{11}^{-1} S_{12} \end{bmatrix} < 0 \Leftrightarrow (ii)\text{成立}.$$

(i)⇔(iii) 类似于(i)⇔(ii)的证明. □

定理 11.1.1 令

$$N = \begin{bmatrix} P & X \\ Y & Z \end{bmatrix},$$

其中X, Y, P和Z是具有适当维数的矩阵，矩阵Z可逆且$N + N^T < 0$，则

$$P + P^T - XZ^{-1}Y - Y^T Z^{-T} X^T < 0. \tag{11.1.1}$$

证明 由Schur补引理及$N + N^T < 0$得到$Z + Z^T < 0$且

$$P + P^T - (X + Y^T)(Z + Z^T)^{-1}(Y + X^T) < 0,$$

即

$$P + P^T - \begin{bmatrix} X^T \\ Y \end{bmatrix}^T \begin{bmatrix} (Z + Z^T)^{-1} & (Z + Z^T)^{-1} \\ (Z + Z^T)^{-1} & (Z + Z^T)^{-1} \end{bmatrix} \begin{bmatrix} X^T \\ Y \end{bmatrix} < 0.$$

既然式(11.1.1)能被写做

$$P + P^T - \begin{bmatrix} X^T \\ Y \end{bmatrix}^T \begin{bmatrix} 0 & Z^{-1} \\ Z^{-T} & 0 \end{bmatrix} \begin{bmatrix} X^T \\ Y \end{bmatrix} < 0,$$

只需证

$$\begin{bmatrix} (Z+Z^T)^{-1} & (Z+Z^T)^{-1} - Z^{-1} \\ (Z+Z^T)^{-1} - Z^{-T} & (Z+Z^T)^{-1} \end{bmatrix} \leqslant 0.$$

事实上, 由Schur补引理及$Z + Z^T < 0$, 只需证

$$(Z+Z^T)^{-1} - [(Z+Z^T)^{-1} - Z^{-1}](Z+Z^T)[(Z+Z^T)^{-1} - Z^{-T}] \leqslant 0.$$

通过直接计算知上式成立. □

定理 11.1.2 设A是实方阵, T是实对称阵, 则存在矩阵$P > 0$使得

$$A^TPA - P + T < 0$$

的充要条件是存在矩阵$X > 0$使得

$$\begin{bmatrix} -X & AX \\ XA^T & -X + XTX \end{bmatrix} < 0.$$

证明 由引理11.1.1知, 存在矩阵$P > 0$使得$A^TPA - P + T < 0$的充要条件是

$$\begin{bmatrix} -P^{-1} & A \\ A^T & -P + T \end{bmatrix} < 0. \tag{11.1.2}$$

对上式左边的矩阵分别左乘和右乘矩阵$\begin{bmatrix} I & 0 \\ 0 & P^{-1} \end{bmatrix}$可得

$$\begin{bmatrix} -P^{-1} & AP^{-1} \\ P^{-1}A^T & -P^{-1} + P^{-1}TP^{-1} \end{bmatrix} < 0.$$

取$X = P^{-1}$, 得证. □

定理 11.1.3 设X和Y是n阶实对称正定矩阵, n_k是正整数, 则下面表述等价:

(i) 存在矩阵$X_2, Y_2 \in \mathbb{R}^{n \times n_k}$和对称正定矩阵$X_3, Y_3 \in \mathbb{R}^{n_k \times n_k}$满足

$$\begin{bmatrix} X & X_2 \\ X_2^T & X_3 \end{bmatrix} > 0, \tag{11.1.3}$$

$$\begin{bmatrix} X & X_2 \\ X_2^T & X_3 \end{bmatrix}^{-1} = \begin{bmatrix} Y & Y_2 \\ Y_2^T & Y_3 \end{bmatrix}. \tag{11.1.4}$$

(ii) $\begin{bmatrix} X & I \\ I & Y \end{bmatrix} \geqslant 0$, $\operatorname{rank}\begin{bmatrix} X & I \\ I & Y \end{bmatrix} \leqslant n + n_k$.

证明 (i)⇒(ii) 由式(11.1.3)及式(11.1.4)得

$$0 \leqslant \begin{bmatrix} I & 0 \\ Y & Y_2 \end{bmatrix}\begin{bmatrix} X & X_2 \\ X_2^T & X_3 \end{bmatrix}\begin{bmatrix} I & Y \\ 0 & Y_2^T \end{bmatrix} = \begin{bmatrix} X & I \\ I & Y \end{bmatrix}.$$

使用

$$\begin{bmatrix} X & I \\ I & Y \end{bmatrix} = \begin{bmatrix} I & Y^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} X - Y^{-1} & 0 \\ 0 & Y \end{bmatrix}\begin{bmatrix} I & 0 \\ Y^{-1} & I \end{bmatrix} \tag{11.1.5}$$

推出

$$\begin{aligned} \operatorname{rank}\begin{bmatrix} X & I \\ I & Y \end{bmatrix} &= n + \operatorname{rank}(X - Y^{-1}) \\ &= n + \operatorname{rank}(XY - I) \\ &= n + \operatorname{rank}(X_2Y_2^T) \leq n + n_k. \end{aligned}$$

(ii)⇒(i) 由(ii)及式(11.1.5)得$X - Y^{-1} \geqslant 0, \operatorname{rank}(X - Y^{-1}) \leqslant n_k$. 因此, 存在矩阵$X_2 \in \mathbb{R}^{n\times n_k}$使得

$$0 \leqslant X - Y^{-1} = X_2X_2^T, \tag{11.1.6}$$

于是$X - X_2X_2^T > 0$. 由引理11.1.1可得

$$\begin{bmatrix} X & X_2 \\ X_2^T & I \end{bmatrix} > 0.$$

另一方面, 由式(11.1.6)得

$$\begin{bmatrix} X & X_2 \\ X_2^T & I \end{bmatrix}^{-1} = \begin{bmatrix} Y & -YX_2 \\ -X_2^TY & X_2^TYX_2 + I \end{bmatrix},$$

故取$X_3 = I, Y_3 = X_2^TYX_2 + I, Y_2 = -YX_2$即可. □

11.2 Projection引理及其应用

本节首先介绍Projection引理, 并由此给出Fisher引理和Elimination引理. 这些引理可用来减少线性矩阵不等式中的决策变量个数, 从而降低计算的复杂性.

引理 11.2.1 设Z是一个实对称矩阵, 且有形式

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{12}^T & Z_{22} & Z_{23} \\ Z_{13}^T & Z_{23}^T & Z_{33} \end{bmatrix},$$

则存在矩阵X使得

$$\begin{bmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{12}^T & Z_{22} & Z_{23}+X \\ Z_{13}^T & Z_{23}^T+X^T & Z_{33} \end{bmatrix} < 0 \tag{11.2.1}$$

的充要条件是

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{12}^T & Z_{22} \end{bmatrix} < 0, \begin{bmatrix} Z_{11} & Z_{13} \\ Z_{13}^T & Z_{33} \end{bmatrix} < 0. \tag{11.2.2}$$

进而, 如果式(11.2.2)成立, 则使得不等式(11.2.1)成立的一个矩阵X可取为

$$X = Z_{12}^T Z_{11}^{-1} Z_{13} - Z_{23}.$$

证明 必要性显然, 下证充分性.

由式(11.2.2)得$Z_{11} < 0$, $Z_{22} - Z_{12}^T Z_{11}^{-1} Z_{12} < 0$且$Z_{33} - Z_{13}^T Z_{11}^{-1} Z_{13} < 0$. 令

$$X = Z_{12}^T Z_{11}^{-1} Z_{13} - Z_{23},$$

则

$$\begin{aligned} &\begin{bmatrix} Z_{22} & Z_{23}+X \\ Z_{23}^T+X^T & Z_{33} \end{bmatrix} - \begin{bmatrix} Z_{12}^T \\ Z_{13}^T \end{bmatrix} Z_{11}^{-1} \begin{bmatrix} Z_{12} & Z_{13} \end{bmatrix} \\ = &\begin{bmatrix} Z_{22} - Z_{12}^T Z_{11}^{-1} Z_{12} & 0 \\ 0 & Z_{33} - Z_{13}^T Z_{11}^{-1} Z_{13} \end{bmatrix} < 0. \end{aligned} \tag{11.2.3}$$

这和Schur补引理一起推出式(11.2.1)成立. □

对于实矩阵M, 符号N_M表示由$\mathcal{N}(M)$的任意一组基作为列排成的矩阵.

定理 11.2.1 (Projection引理) [16] 设P, Q和H是给定的有适当维数的实矩阵, 且H是对称的, P, Q均非列满秩, 则存在矩阵X使得

$$H + P^T X^T Q + Q^T X P < 0 \tag{11.2.4}$$

的充要条件是

$$N_P^T H N_P < 0, N_Q^T H N_Q < 0. \tag{11.2.5}$$

证明　必要性. 由矩阵N_P和N_Q的定义得$PN_P = 0, QN_Q = 0$. 对式(11.2.4)两端左乘矩阵N_P^T、右乘矩阵N_P得$N_P^T H N_P < 0$. 同理$N_Q^T H N_Q < 0$.

充分性. 设V_1是由$\mathcal{N}(P) \cap \mathcal{N}(Q)$的任意一组基作为列所排成的矩阵, 则存在列满秩矩阵V_2和V_3使得

$$\mathcal{R}\left(\begin{bmatrix} V_1 & V_2 \end{bmatrix}\right) = \mathcal{N}(P), \quad \mathcal{R}\left(\begin{bmatrix} V_1 & V_3 \end{bmatrix}\right) = \mathcal{N}(Q),$$

并且$\begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix}$的各列构成$\mathcal{N}(P) + \mathcal{N}(Q)$的一组基. 因此, 存在矩阵$V_4$使得$V = \begin{bmatrix} V_1 & V_2 & V_3 & V_4 \end{bmatrix}$可逆, 故矩阵不等式(11.2.4)成立的充要条件是

$$V^T H V + V^T P^T X^T Q V + V^T Q^T X P V < 0. \tag{11.2.6}$$

根据矩阵V的构造有

$$PV = \begin{bmatrix} 0 & 0 & P_1 & P_2 \end{bmatrix}, \ QV = \begin{bmatrix} 0 & Q_1 & 0 & Q_2 \end{bmatrix}.$$

设

$$V^T H V = \begin{bmatrix} H_{11} & H_{12} & H_{13} & H_{14} \\ H_{12}^T & H_{22} & H_{23} & H_{24} \\ H_{13}^T & H_{23}^T & H_{33} & H_{34} \\ H_{14}^T & H_{24}^T & H_{34}^T & H_{44} \end{bmatrix},$$

$$P_i^T X^T Q_j = Y_{ij}, \quad i, j = 1, 2, \tag{11.2.7}$$

则矩阵不等式(11.2.6)变成

$$\begin{bmatrix} H_{11} & H_{12} & H_{13} & H_{14} \\ H_{12}^T & H_{22} & H_{23} + Y_{11}^T & H_{24} + Y_{21}^T \\ H_{13}^T & H_{23}^T + Y_{11} & H_{33} & H_{34} + Y_{12} \\ H_{14}^T & H_{24}^T + Y_{21} & H_{34}^T + Y_{12}^T & H_{44} + Y_{22} + Y_{22}^T \end{bmatrix} < 0. \tag{11.2.8}$$

注意到$\mathcal{N}\left(\begin{bmatrix} P_1 & P_2 \end{bmatrix}\right) = 0$及$\mathcal{N}\left(\begin{bmatrix} Q_1 & Q_2 \end{bmatrix}\right) = 0$, 对任意给定的矩阵$Y_{i,j}$, $i, j = 1, 2$, 存在矩阵X使得式(11.2.7)成立. 故只需证存在矩阵Y_{ij}, $i, j = 1, 2$使得式(11.2.8)成立. 由Schur补引理得式(11.2.8)成立的充要条件是以下的两个矩阵不等式成立:

$$\overline{H} := \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{12}^T & H_{22} & H_{23} + Y_{11}^T \\ H_{13}^T & H_{23}^T + Y_{11} & H_{33} \end{bmatrix} < 0, \tag{11.2.9}$$

$$H_{44} + Y_{22} + Y_{22}^T - \begin{bmatrix} H_{14} \\ H_{24} + Y_{21}^T \\ H_{34} + Y_{12} \end{bmatrix}^T \overline{H}^{-1} \begin{bmatrix} H_{14} \\ H_{24} + Y_{21}^T \\ H_{34} + Y_{12} \end{bmatrix} < 0. \tag{11.2.10}$$

如果能选取一个适当的Y_{11}使得不等式(11.2.9)成立，则总可以选取适当的Y_{12}, Y_{21}和Y_{22} 使得不等式(11.2.10)成立. 因此，存在矩阵Y_{ij}, $i,j=1,2$使得不等式(11.2.8)成立的充要条件是存在矩阵Y_{11}使得式(11.2.9)成立.

另一方面，由

$$\mathcal{N}(PV)=\mathcal{R}\left(\begin{bmatrix} I & 0 \\ 0 & I \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\right), \mathcal{N}(QV)=\mathcal{R}\left(\begin{bmatrix} I & 0 \\ 0 & 0 \\ 0 & I \\ 0 & 0 \end{bmatrix}\right)$$

及式(11.2.5)推出

$$\begin{bmatrix} H_{11} & H_{12} \\ H_{12}^T & H_{22} \end{bmatrix} < 0, \quad \begin{bmatrix} H_{11} & H_{13} \\ H_{13}^T & H_{33} \end{bmatrix} < 0.$$

应用引理11.2.1，存在矩阵Y_{11}使得$\overline{H}<0$，即式(11.2.9)成立，进而存在矩阵X使得不等式(11.2.4)成立. □

定理 11.2.2 (Fisher**引理**) [16, Lemma 3] 设$L^T=L\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{m\times n}$, $\text{rank}B<n$, N_B同前面定义，则下列表述等价:

(i) $p^TLp<0,\ \forall 0\neq p\in\mathcal{N}(B)$;

(ii) $(N_B)^TLN_B<0$;

(iii) 存在$l>0$使得$L-lB^TB<0$;

(iv) 存在$X\in\mathbb{R}^{n\times m}$使得$L+XB+B^TX^T<0$.

证明 (i)⇒(ii). 设B的奇异值分解为

$$B=U\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}V, \tag{11.2.11}$$

其中U和V为正交阵，Σ为$r\times r$的正定对角阵，于是

$$\mathcal{N}(B)=\left\{V^T\begin{bmatrix} 0 \\ \xi \end{bmatrix}: \xi\in\mathbb{R}^{n-r}\right\},$$

从而

$$N_B=V^T\begin{bmatrix} 0 \\ B_1 \end{bmatrix}, \tag{11.2.12}$$

其中$B_1\in\mathbb{R}^{(n-r)\times(n-r)}$可逆. 设

$$L=V^T\begin{bmatrix} L_{11} & L_{12} \\ L_{12}^T & L_{22} \end{bmatrix}V, \tag{11.2.13}$$

则由(i)知$L_{22}<0$, 故

$$(N_B)^T L N_B = \begin{bmatrix} 0 & B_1^T \end{bmatrix} VV^T \begin{bmatrix} L_{11} & L_{12} \\ L_{12}^T & L_{22} \end{bmatrix} VV^T \begin{bmatrix} 0 \\ B_1 \end{bmatrix} = B_1^T L_{22} B_1 < 0,$$

即(ii)成立.

(ii)⇒(iii). 设B, N_B及L分别有形式(11.2.11), (11.2.12)及(11.2.13), 则

$$L - lB^T B = V^T \begin{bmatrix} L_{11} - l\Sigma^2 & L_{12} \\ L_{12}^T & L_{22} \end{bmatrix} V,$$

$$N_B^T L N_B = B_1^T L_{22} B_1.$$

由(ii)知$L_{22}<0$, 故存在$l>0$使得$L-lB^TB<0$的充要条件是存在$l>0$使得

$$L_{11} - l\Sigma^2 - L_{12}L_{22}^{-1}L_{12}^T < 0.$$

取l充分大即可.

(iii)⇒(iv) 取$X=-\frac{l}{2}B^T$即可.

(iv)⇒(i) 显然. □

注记 11.2.1 定理11.2.2中(iv)与(ii)的等价性可看做是定理11.2.1的特殊情况.

引理 11.2.2 设P为n阶实对称矩阵, $\mathbb{U}$是$\mathbb{R}^n$的k维非零子空间. 如果

$$x^T P x < 0, \quad \forall x \in \mathbb{U} \setminus \{0\},$$

则P至少有k个负特征值.

证明 设P有t个负特征值, 则存在可逆阵C使得

$$P = C^T \operatorname{diag}(-I_t, I_s, 0) C.$$

令

$$\mathbb{V} = \mathcal{R}\left(C^{-1}\begin{bmatrix} 0 \\ I_{n-t} \end{bmatrix}\right),$$

则$\dim\mathbb{V}=n-t$, 并且$y^TPy\geqslant 0$, $\forall y\in\mathbb{V}$. 这和已知一起推出$\mathbb{U}\cap\mathbb{V}=\{0\}$, 于是$\dim\mathbb{U}+\dim\mathbb{V}=\dim(\mathbb{U}+\mathbb{V})\leq n$, 从而$k=\dim\mathbb{U}\leqslant t$, 所以$P$至少有$k$个负特征值. □

下面的引理可类似于引理11.2.2证得(留给读者).

引理 11.2.3 设P为n阶实对称可逆阵, $\mathbb{U}$是$\mathbb{R}^n$的k维非零子空间. 如果

$$x^T P x \leqslant 0, \ \forall x \in \mathbb{U},$$

则P至少有k个负特征值.

定理 11.2.3 (Elimination引理) 设

$$P = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}, \ P^{-1} = \begin{bmatrix} \tilde{Q} & \tilde{S} \\ \tilde{S}^T & \tilde{R} \end{bmatrix},$$

其中$R \geqslant 0$, $\tilde{Q} \leqslant 0$且Q和$\tilde{Q}$的阶数相同, 则存在矩阵X满足

$$\begin{bmatrix} I \\ A^T X B + C \end{bmatrix}^T P \begin{bmatrix} I \\ A^T X B + C \end{bmatrix} < 0 \tag{11.2.14}$$

的充要条件是

$$N_B^T \begin{bmatrix} I \\ C \end{bmatrix}^T P \begin{bmatrix} I \\ C \end{bmatrix} N_B < 0 \tag{11.2.15}$$

且

$$N_A^T \begin{bmatrix} -C^T \\ I \end{bmatrix}^T P^{-1} \begin{bmatrix} -C^T \\ I \end{bmatrix} N_A > 0. \tag{11.2.16}$$

证明 不妨设$R > 0$(否则, 用$R + \varepsilon I$代替R即可, 其中ε为可任意小的正数).

通过直接计算得到

$$\begin{aligned} & \begin{bmatrix} I \\ A^T X B + C \end{bmatrix}^T P \begin{bmatrix} I \\ A^T X B + C \end{bmatrix} \\ = & \begin{bmatrix} I \\ C \end{bmatrix}^T P \begin{bmatrix} I \\ C \end{bmatrix} + (A^T X B)^T (S^T + RC) \\ & + (S^T + RC)^T (A^T X B) + (A^T X B)^T R (A^T X B). \end{aligned}$$

根据Schur补引理, 式(11.2.14)等价于

$$\begin{bmatrix} \begin{bmatrix} I \\ C \end{bmatrix}^T P \begin{bmatrix} I \\ C \end{bmatrix} & 0 \\ 0 & -R^{-1} \end{bmatrix} + \begin{bmatrix} (S^T + RC)^T A^T \\ A^T \end{bmatrix} X \begin{bmatrix} B & 0 \end{bmatrix}$$
$$+ \begin{bmatrix} B^T \\ 0 \end{bmatrix} X^T \begin{bmatrix} A(S^T + RC) & A \end{bmatrix} < 0. \tag{11.2.17}$$

使用Projection引理, 存在矩阵X满足式(11.2.14)的充要条件是式(11.2.15)和下面的不等式同时成立:

$$\begin{bmatrix} I & 0 \\ -(S^T+RC) & N_A \end{bmatrix}^T \begin{bmatrix} \begin{bmatrix} I \\ C \end{bmatrix}^T P \begin{bmatrix} I \\ C \end{bmatrix} & 0 \\ 0 & -R^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -(S^T+RC) & N_A \end{bmatrix} < 0. \tag{11.2.18}$$

因而只需证(11.2.16)⇔(11.2.18). 事实上, 由P的形式, 不等式(11.2.18)能够变形为

$$\begin{bmatrix} Q-SR^{-1}S^T & (SR^{-1}+C^T)N_A \\ N_A^T(R^{-1}S^T+C) & -N_A^TR^{-1}N_A \end{bmatrix} < 0.$$

由已知及引理11.2.3得P的负特征值个数为Q的阶数、正特征值个数为R的阶数, 于是$Q-SR^{-1}S^T<0$, 这和Schur补引理一起推出式(11.2.18)成立的充要条件是

$$\begin{aligned} & N_A^T \begin{bmatrix} -C^T \\ I \end{bmatrix}^T P^{-1} \begin{bmatrix} -C^T \\ I \end{bmatrix} N_A \\ = \; & N_A^T[R^{-1}+(R^{-1}S^T+C)(Q-SR^{-1}S^T)^{-1}(SR^{-1}+C^T)]N_A > 0 \end{aligned}$$

等价于

$$\begin{bmatrix} N_A^TR^{-1}N_A & N_A^T(R^{-1}S^T+C) \\ (SR^{-1}+C^T)N_A & SR^{-1}S^T-Q \end{bmatrix} > 0,$$

即(11.2.16)⇔(11.2.18). □

由定理11.2.3的证明易得下面的推论:

推论 11.2.1 设

$$P = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix},$$

其中$R>0$, $Q-SR^{-1}S^T<0$, 则存在矩阵X满足

$$\begin{bmatrix} I \\ A^TXB+C \end{bmatrix}^T P \begin{bmatrix} I \\ A^TXB+C \end{bmatrix} < 0 \tag{11.2.19}$$

的充要条件是

$$N_B^T \begin{bmatrix} I \\ C \end{bmatrix}^T P \begin{bmatrix} I \\ C \end{bmatrix} N_B < 0 \tag{11.2.20}$$

且

$$N_A^T \begin{bmatrix} -C^T \\ I \end{bmatrix}^T P^{-1} \begin{bmatrix} -C^T \\ I \end{bmatrix} N_A > 0. \tag{11.2.21}$$

11.3 Dualization引理

定理 11.3.1 (Dualization**引理**) 设P是n阶实对称可逆阵, $\mathbb{R}^n$的两个子空间$\mathcal{U}$和$\mathcal{V}$满足$\mathcal{U}\oplus\mathcal{V}=\mathbb{R}^n$, 则

$$x^TPx<0,\forall x\in\mathcal{U}\setminus\{0\} \text{ 且 } x^TPx\geqslant 0,\forall x\in\mathcal{V}$$

的充要条件是

$$x^TP^{-1}x>0,\forall x\in\mathcal{U}^{\perp}\setminus\{0\} \text{ 且 } x^TP^{-1}x\leqslant 0,\forall x\in\mathcal{V}^{\perp}.$$

证明 因为$\mathcal{U}\oplus\mathcal{V}=\mathbb{R}^n$与$\mathcal{U}^{\perp}\oplus\mathcal{V}^{\perp}=\mathbb{R}^n$等价, 只需证必要性.

设$\dim\mathcal{U}=k$, $\dim\mathcal{V}=l$, 则由已知条件和引理11.2.2和11.2.3知P恰好有k个负特征值、l个正特征值.

首先证明: $x^TP^{-1}x>0,\forall x\in\mathcal{U}^{\perp}\setminus\{0\}$. 反证法. 如果存在$y\in\mathcal{U}^{\perp}\setminus\{0\}$使得$y^TP^{-1}y\leqslant 0$. 令$z=P^{-1}y$. 若$z\in\mathcal{U}$, 则$y^Tz=0$, 于是$z^TPz=0$, 与已知矛盾, 故$z\notin\mathcal{U}$. 因而线性空间$\mathcal{U}_e:=\operatorname{span}\{z\}+\mathcal{U}$的维数为$k+1$. 任取$w\in\mathcal{U}_e$, 则存在$x\in\mathcal{U}$和$k\in\mathbb{R}$使得$w=x+kz$, 于是

$$\begin{aligned} w^TPw= &\ (x+kz)^TP(x+kz)=k^2y^TP^{-1}y+ky^Tx+kx^Ty+x^TPx\\ = &\ k^2y^TP^{-1}y+x^TPx\leq 0. \end{aligned}$$

使用引理11.2.3推出P至少有$k+1$个负特征值, 这和P恰好有k个负特征值矛盾.

下面证明: $x^TP^{-1}x\geqslant 0,\forall x\in\mathcal{V}^{\perp}$. 事实上, 对于任意小的$\varepsilon>0$, 由已知条件知

$$x^T(P+\varepsilon I)x>0,\forall x\in\mathcal{V}\setminus\{0\}.$$

类似于前面的证明得到

$$x^T(P+\varepsilon I)^{-1}x<0,\forall x\in\mathcal{V}^{\perp}\setminus\{0\},$$

于是

$$x^TP^{-1}x=\lim_{\varepsilon\to 0}x^T(P+\varepsilon I)^{-1}x\leqslant 0,\forall x\in\mathcal{V}^{\perp}.$$

□

11.4 含线性参数的线性矩阵不等式

本节介绍几个含线性参数的线性矩阵不等式, 它们的共同特点是参数在给定的紧集中变化. 这些不等式可应用于线性参数变化系统的鲁棒控制问题.

引理 11.4.1 [17] 设A, B为有合适维数的实矩阵, $A = A^T$, $\mathbb{H}$是包含一些实对称矩阵的紧集, 则下列命题等价:

(i) $A < BXB^T$, $\forall X \in \mathbb{H}$;

(ii) 存在实对称阵Y使得$A < BYB^T, Y < X$, $\forall X \in \mathbb{H}$.

证明 (ii)⇒(i). 显然, 下面证明(i)⇒(ii). 设B有等价分解

$$B = M \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} N,$$

其中M, N为可逆阵, $r = \mathrm{rank}B$. 记

$$A = M \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} M^T,\ X = N^{-1} \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} N^{-T},$$

则(i)等价于

$$A_{22} < 0,\ A_{11} - A_{12}A_{22}^{-1}A_{12}^T < X_{11},\ \forall X \in \mathbb{H}. \tag{11.4.1}$$

将X_{11}看成关于X的函数$X_{11} = f(X)$, 并定义集合

$$\tilde{\mathbb{H}} = \{f(X) : x \in \mathbb{H}\},$$

由$\mathbb{H}$是紧集知$\tilde{\mathbb{H}}$是紧集. 令

$$\lambda_1 = \min_{Z \in \tilde{\mathbb{H}}} \lambda_{\min}(Z),\ \lambda_2 = \lambda_{\max}(A_{11} - A_{12}A_{22}^{-1}A_{12}^T),$$

则由式(11.4.1)易见$\lambda_2 < \lambda_1$. 取充分小的ε(满足$0 < \varepsilon < \lambda_1 - \lambda_2$), 并设

$$Y = N^{-1} \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{12}^T + \varepsilon I & 0 \\ 0 & -\frac{1}{\varepsilon} I \end{bmatrix} N^{-T},$$

则由Schur补引理易见

$$A < BYB^T,\ Y < X,\ \forall X \in \mathbb{H},$$

即(ii)成立. □

引理 11.4.2 设$Q \in \mathbb{R}^{n\times n}$, $F \in \mathbb{R}^{n\times p}$, $Q^T = Q$, $\mathbb{H}$是实$p \times p$对称矩阵空间中的紧子集. 若对于所有的$H \in \mathbb{H}$均有$HF \neq 0$且

$$\xi^T Q \xi < 0,\ \forall 0 \neq \xi \in \mathcal{N}(HF),$$

则存在正数Γ使得

$$Q < \Gamma F^T H^T HF,\ \forall H \in \mathbb{H}.$$

证明 设F的等价分解为$F = U\text{diag}(I,0)V$, 于是$HF = \begin{bmatrix}\hat{H} & 0\end{bmatrix} V$. 由已知条件知$\hat{H} \neq 0$, 从而可设

$$HF = U_1 \begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix} V_1,$$

其中U_1是正交阵, V_1是可逆阵, $q > 0$ (注意: U_1, V_1, q依赖于H). 对应地, 设

$$Q = V_1^T \begin{bmatrix} Q_1 & Q_2 \\ Q_2^T & Q_3 \end{bmatrix} V_1, \; Q_1 \in \mathbb{R}^{q\times q},$$

则由已知条件知$Q_3 < 0$, 并由矩阵的合同知, 只需证明: 存在正数Γ使得

$$Q_1 - Q_2 Q_3^{-1} Q_2^T - \Gamma I_q < 0.$$

设$\|.\|$是矩阵范数, 由上面变换过程知, 范数$\|Q_1 - Q_2Q_3^{-1}Q_2^T\|$关于$H$中的元素是连续函数. 又由$\mathbb{H}$是紧集知$\mathbb{H}$是有界闭集, 则$\|Q_1 - Q_2Q_3^{-1}Q_2^T\|$在$\mathbb{H}$上有最大值$\mu_{\max}$. 所以矩阵的谱半径满足

$$\rho(Q_1 - Q_2Q_3^{-1}Q_2^T) \leqslant \mu_{\max},$$

这样选取$\Gamma > \mu_{\max}$即可. □

定理 11.4.1 [17] 设$Q \in \mathbb{R}^{n\times n}, F \in \mathbb{R}^{n\times p}$, $Q = Q^T$, $\mathbb{H}$是实$p\times p$对称矩阵空间中的紧子集. 则下列命题等价:

(i) 对于所有的$H \in \mathbb{H}$有

$$\xi^T Q \xi < 0, \; \forall 0 \neq \xi \in \mathcal{N}(HF).$$

(ii) 存在实对称阵Z使得

$$Q + F^T Z F < 0, \; N_H^T Z N_H \geqslant 0, \; \forall H \in \mathbb{H}.$$

证明 (i)⇒(ii). 若存在$H \in \mathbb{H}$使得$HF = 0$, 则$\mathcal{N}(HF) = \mathbb{R}^p$, 从而$Q < 0$, 故$Z = 0$满足(ii).

若$HF = 0, \forall H \in \mathbb{H}$, 由引理11.4.2知存在$\Gamma_0 > 0$使得

$$\lambda_{\min}(F^T H^T H F) > \Gamma_0, \; \forall H \in \mathbb{H},$$

从而

$$Q < \Gamma F^T H^T H F, \; \forall H \in \mathbb{H}.$$

令$\mathbb{H}_1 = \left\{\Gamma H^T H : H \in \mathbb{H}\right\}$, 则由$\mathbb{H}$是紧集知$\mathbb{H}_1$是紧集. 故上式等价于

$$Q < F^T S F, \; \forall S \in \mathbb{H}_1.$$

应用引理11.4.1, 存在实对称阵Z使得

$$Q + F^T Z F < 0,\ -Z < \Gamma H^T H,\ \forall H \in \mathbb{H},$$

从而(ii)成立.

(ii)⇒(i). 固定$H \in \mathbb{H}$, 任取$0 \neq \xi \in \mathcal{N}(HF)$, 则$F\xi \in \mathcal{N}(H)$, 于是存在$\eta$使得$F\xi = N_H \eta$, 故

$$\xi^T Q \xi < -\xi^T F^T Z F \xi = -\eta^T N_H^T Z N_H \eta \leq 0.$$

□

命题 11.4.1 [18] 考虑一个标量二次函数

$$f(\delta_1, \cdots, \delta_q) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \delta_i + \sum_{i<j} \beta_{ij} \delta_i \delta_j + \sum_{i=1}^{q} \gamma_i \delta_i^2,$$

若$f(\cdot)$是多凸的, 即

$$\gamma_i := \frac{\partial^2}{\partial \delta_i^2} f \geqslant 0,\ i = 1, 2, \cdots, q,$$

则$f(\cdot) < 0$在超长方体

$$H := \left\{(\delta_1, \cdots, \delta_q) : \delta_i \in [\underline{\delta_i}, \overline{\delta_i}], i = 1, 2, \cdots, q\right\}$$

中任一点成立的充要条件是$f(\cdot) < 0$在H的顶点集V上成立.

证明 必要性显然, 下证充分性.

设$\delta^* = (\delta_1^*, \cdots, \delta_q^*)$是$f(\cdot)$在$H$上的一个全局最大值点, 考虑函数

$$g(\delta_i) = f(\delta_1^*, \cdots, \delta_{i-1}^*, \delta_i, \delta_{i+1}^*, \cdots, \delta_q^*).$$

注意到$g(\delta_i)$有形式

$$g(\delta_i) = a + b\delta_i + \gamma_i \delta_i^2,$$

由$f(\cdot)$的多凸性推出$g(\delta_i)$是一个凸函数. 根据凸函数的性质, $g(\delta_i)$在区间$[\underline{\delta_i}, \overline{\delta_i}]$上的最大值必在区间端点$\underline{\delta_i}$和$\overline{\delta_i}$中的一个达到, 从而

$$g(\delta_i^*) \leqslant \max\{g(\underline{\delta_i}), g(\overline{\delta_i})\}.$$

另一方面, 由于δ^*是$f(\cdot)$的一个全局最大值点, 故

$$\max\{g(\underline{\delta_i}), g(\overline{\delta_i})\} \leqslant f(\delta^*) = g(\delta_i^*).$$

组合这两个不等式得到

$$g(\delta_i^*) = \max\{g(\underline{\delta_i}), g(\overline{\delta_i})\}.$$

重复这样的讨论, 可得$f(\cdot)$可以在H的顶点集V上达到最大值. 从而结论得证.
□

推论 11.4.1 [19] 设$a, b \in \mathbb{R}$满足$a \leqslant b$, X_1, X_2和Y是给定的n阶实对称矩阵, 则

$$(f-a)X_1 + (b-f)X_2 + Y < 0,\ \forall f \in [a, b]$$

的充要条件是

$$(b-a)X_1 + Y < 0,\ (b-a)X_2 + Y < 0.$$

证明 任取$z \in \mathbb{R}^n$, 由命题11.4.1知

$$\begin{aligned} & z^T[(f-a)X_1 + (b-f)X_2 + Y]z < 0 \\ \Leftrightarrow\ & z^T[(b-a)X_1 + Y]z < 0 \text{ 且 } z^T[(f-a)X_2 + Y]z < 0, \end{aligned}$$

从而结论成立. □

11.5 鲁棒控制中的几个基础不等式

本节介绍鲁棒控制中的几个基本的线性矩阵不等式, 尽管这些不等式不经常直接用来处理问题, 但鲁棒控制中的几个重要不等式是由这些不等式推出的.

首先给出下面的定理, 证明见附录F.

定理 11.5.1 设X, Y和Z均为n阶实对称阵, $T > 0$,

$$\Phi(\lambda) = \lambda^2 X + \lambda Y + Z, \forall \lambda \in \mathbb{R}.$$

若$\lambda_{\max}(\Phi(\lambda)) > 0$, $\forall \lambda \in [0, T]$, 则存在连续函数$\eta : [0, T] \to \mathbb{R}^n$使得

(i) $\eta^T(\lambda)\Phi(\lambda)\eta(\lambda) > 0, \quad \forall \lambda \in [0, T]$;

(ii) $\Phi(0)\eta(0) = \lambda_{\max}(\Phi(0))\eta(0), \quad \Phi(T)\eta(T) = \lambda_{\max}(\Phi(T))\eta(T)$.

定理 11.5.2 设n阶实对称阵X, Y和Z满足$X > 0$, 若

(i) $\delta(x) := (x^T Y x)^2 - 4(x^T X x)(x^T Z x) > 0, \forall 0 \neq x \in \mathbb{R}^n$;

(ii) $x^T Y x < 0$, $\forall x \in S$, 其中$S = \{x \in \mathbb{R}^n | x \neq 0, x^T Z x \geqslant 0\}$,

则存在$\lambda > 0$使得$\Phi(\lambda) := \lambda^2 X + \lambda Y + Z \leq 0$.

证明 不妨设$X = I_n$.

(a) 当$Z \leqslant 0$时, 不妨设

$$Z = \text{diag}(Z_1, 0), \quad Z_1 < 0.$$

由(ii)知

$$Y = \begin{bmatrix} Y_1 & Y_2 \\ Y_2^T & Y_3 \end{bmatrix}, \quad Y_3 < 0,$$

于是$\Phi(\lambda)$合同于

$$\begin{bmatrix} \lambda^2 I + \lambda Y_1 + Z_1 + \lambda Y_2(\lambda I + Y_3)^{-1} Y_2^T & 0 \\ 0 & \lambda^2 I + \lambda Y_3 \end{bmatrix},$$

故当λ充分小时结论成立.

(b) 当$Z \nleqslant 0$时, 则$\lambda_{\max}(Z) > 0$. 若结论不成立, 则$\lambda_{\max}(\Phi(\lambda)) > 0, \forall \lambda \geqslant 0$. 既然当$\lambda$充分大时$\Phi(\lambda) > 0$, 使用定理11.5.1知存在连续函数$\eta : [0, +\infty) \to \mathbb{R}^n$满足

$$\eta^T(\lambda)\eta(\lambda) = 1 \text{ 且 } \eta^T(\lambda)\Phi(\lambda)\eta(\lambda) > 0, \forall \lambda \geqslant 0,$$

于是

$$\lambda^2 + \lambda\eta^T(\lambda)Y\eta(\lambda) + \eta^T(\lambda)Z\eta(\lambda) > 0, \forall \lambda \geqslant 0,$$

即

$$(\lambda - \rho_1(\lambda))(\lambda - \rho_2(\lambda)) > 0, \forall \lambda \geqslant 0, \tag{11.5.1}$$

其中

$$\rho_1(\lambda) = \frac{-\eta^T(\lambda)Y\eta(\lambda) - \sqrt{\delta(\eta(\lambda))}}{2}, \ \forall \lambda \geqslant 0,$$

$$\rho_2(\lambda) = \frac{-\eta^T(\lambda)Y\eta(\lambda) + \sqrt{\delta(\eta(\lambda))}}{2}, \ \forall \lambda \geqslant 0.$$

显然

$$\rho_1(\lambda) < \rho_2(\lambda), \forall \lambda \geqslant 0.$$

由式(11.5.1)推出$\lambda > \rho_2(\lambda) > \rho_1(\lambda)$或$\lambda < \rho_1(\lambda) < \rho_2(\lambda), \forall \lambda \geqslant 0$.

若存在λ_0使得

$$\lambda_0 > \rho_2(\lambda_0) > \rho_1(\lambda_0),$$

则由$\lambda - \rho_2(\lambda)$的连续性及式(11.5.1)推出

$$\lambda > \rho_2(\lambda), \forall \lambda \geqslant 0,$$

于是

$$\frac{-\eta^T(0)Y\eta(0) + \sqrt{\delta(\eta(0))}}{2} < 0.$$

故$\eta^T(0)Y\eta(0) > 0$且$\eta^T(0)Z\eta(0) > 0$, 与(ii)矛盾.

若$\lambda < \rho_2(\lambda), \forall \lambda \geqslant 0$, 则$\lambda < \dfrac{-\eta^T(\lambda)Y\eta(\lambda) - \sqrt{\delta(\eta(\lambda))}}{2}, \forall \lambda \geqslant 0$, 矛盾.

综上所述结论成立. □

引理 11.5.1 设正整数n和非负整数p,q满足$p+q\leqslant n$,

$$S=\left\{\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}^T\in\mathbb{R}^n:\sum_{i=1}^{n}x_i^2=1,\ \sum_{i=1}^{p}x_i^2\geqslant\sum_{i=p+1}^{p+q}x_i^2\right\},$$

则S为闭集.

证明 任取S中的收敛子列

$$\begin{bmatrix}x_1^{(i)} & x_2^{(i)} & \cdots & x_n^{(i)}\end{bmatrix}^T, i=1,2,\cdots,$$

设其在$\mathbb{R}^n$中的极限为$\begin{bmatrix}\widehat{x}_1 & \widehat{x}_2 & \cdots & \widehat{x}_n\end{bmatrix}^T$. 令

$$y^{(i)}=1-(x_1^{(i)})^2-\cdots-(x_n^{(i)})^2,$$

$$z^{(i)}=(x_1^{(i)})^2+\cdots+(x_p^{(i)})^2-(x_{p+1}^{(i)})^2-\cdots-(x_{p+q}^{(i)})^2,$$

则$y^{(i)}=0, z^{(i)}\geqslant 0, i=1,2,\cdots$, 于是

$$1-\widehat{x}_1^2-\cdots-\widehat{x}_n^2=\lim_{i\to\infty}y^{(i)}=0,$$

$$\widehat{x}_1^2+\cdots+\widehat{x}_p^2-\widehat{x}_{p+1}^2-\cdots-\widehat{x}_{p+q}^2=\lim_{i\to\infty}z^{(i)}\geqslant 0,$$

从而$\begin{bmatrix}\widehat{x}_1 & \widehat{x}_2 & \cdots & \widehat{x}_n\end{bmatrix}^T\in S$,故$S$是闭集. □

推论 11.5.1 设M为n阶实对称阵, $S=\{x\in\mathbb{R}^n|x^TMx\geqslant 0, x^Tx=1\}$, 则$S$是闭集.

定理 11.5.3 设n阶实对称阵X, Y和Z满足$X\geqslant 0$, 若

(i) $x^TYx<0, \forall x\in S$, 其中$S=\{x\in\mathbb{R}^n|x\neq 0, x^TZx\geqslant 0\}$;

(ii) $\delta(x):=(x^TYx)^2-4(x^TXx)(x^TZx)>0, \forall 0\neq x\in\mathbb{R}^n$,

则存在$\lambda>0$使得

$$\Phi(\lambda):=\lambda^2X+\lambda Y+Z<0. \tag{11.5.2}$$

证明 若$Z<0$, 即$\lambda_{\max}(Z)<0$, 此时容易证明总能找到适当的$\lambda>0$(充分小)使得式(11.5.2)成立.

以下考虑$\lambda_{\max}(Z)\geqslant 0$的情形, 往证: 存在一个常数$\varepsilon>0$使得对称矩阵$\overline{X}:=X+\varepsilon I$, Y和Z满足定理11.5.2的所有条件. 事实上, 由条件(ii)可得

$$\mu=\min\{\delta(x):\|x\|_2=1, x^TZx\geqslant 0\}>0, \tag{11.5.3}$$

选取$\varepsilon > 0$使得

$$4\lambda_{\max}(Z)\varepsilon < \mu. \tag{11.5.4}$$

考虑

$$\overline{\delta}(x) = (x^T Y x)^2 - 4(x^T \overline{X} x)(x^T Z x), \tag{11.5.5}$$

由于$X \geqslant 0$, $\varepsilon > 0$, 故

$$\overline{X} = X + \varepsilon I > 0.$$

因此, 对所有使得$x^T Z x < 0$的非零向量$x \in \mathbb{R}^n$, $\overline{\delta}(x) > 0$. 另一方面, 从式(11.5.3)至式(11.5.5), 可知对使得$x^T Z x \geqslant 0$的所有单位向量x,

$$\overline{\delta}(x) = \delta(x) - 4\varepsilon x^T Z x \geqslant \delta(x) - 4\varepsilon\lambda_{\max}(Z) \geqslant (\mu - 4\varepsilon\lambda_{\max}(Z)) > 0,$$

由此推出定理11.5.2的条件满足, 故根据定理11.5.2存在常数$\lambda > 0$使得

$$\lambda^2(X + \varepsilon I) + \lambda Y + Z \leqslant 0,$$

从而

$$\lambda^2 X + \lambda Y + Z \leqslant -\varepsilon\lambda^2 I < 0.$$

□

引理 11.5.2 设Y和Z均为n阶实对称阵, $Z \leqslant 0$,

$$S_0 = \{\xi \in \mathbb{R}^n | \xi^T Z \xi = 0, \xi \neq 0\},$$

则$\xi^T Y \xi < 0, \forall \xi \in S_0$的充要条件是存在$\lambda > 0$使得$\lambda Y + Z < 0$.

证明 充分性显然, 下证必要性.

若$Z < 0$,则取λ为充分小的正数即可. 若$Z \nless 0$, 则由$Z \leqslant 0$不妨设

$$Z = \mathrm{diag}(Z_1, 0),$$

其中$Z_1 < 0$, $Z_1 \in \mathbb{R}^{r\times r}$, 于是$S_0 = \left\{\begin{bmatrix} 0 & \eta \end{bmatrix}^T : 0 \neq \eta \in \mathbb{R}^{n-r}\right\}$, 从而

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}, Y_{22} < 0,$$

故$\lambda Y + Z$合同于

$$\begin{bmatrix} \lambda(Y_{11} - Y_{12}Y_{22}^{-1}Y_{12}^T) + Z_1 & 0 \\ 0 & \lambda Y_{22} \end{bmatrix}.$$

故存在$\lambda_0 > 0$(充分小)使得$\lambda_0 Y + Z < 0$. □

定理 11.5.4 设Y和Z均为n阶实对称阵,

$$S = \{\xi \in \mathbb{R}^n | \xi^T Z \xi \geqslant 0, \xi \neq 0\},$$

则$\xi^T Y \xi < 0, \forall \xi \in S$的充要条件是存在$\lambda > 0$使得$\lambda Y + Z < 0$.

证明 充分性显然, 下证必要性.

当$Z \leqslant 0$时, 由引理11.5.2得证. 当$Z \nleqslant 0$时, 则$\lambda_{\max}(Z) > 0$. 令

$$S_1 = \{\xi \in S : \xi^T \xi = 1\}, S_2 = \{\xi \in \mathbb{R}^n : \xi^T Z \xi > 0, \xi^T \xi = 1\}.$$

由推论11.5.1知S_1是闭集. 取$0 < \varepsilon < \dfrac{\max_{\xi \in S_1} (\xi^T Y \xi)^2}{4\lambda_{\max}(Z)}$, 则

$$\varepsilon < \frac{\inf\limits_{\xi \in S_2} (\xi^T Y \xi)^2}{4\lambda_{\max}(Z)} = \inf_{\xi \in S_2} \frac{(\xi^T Y \xi)^2}{4\lambda_{\max}(Z)} \leqslant \inf_{\xi \in S_2} \frac{(\xi^T Y \xi)^2}{4\xi^T Z \xi},$$

于是$(\xi^T Y \xi)^2 - 4\varepsilon \xi^T Z \xi > 0, \forall \xi \in S_2$, 故

$$(\eta^T Y \eta)^2 - 4\varepsilon \|\eta\|_2^2 \eta^T Z \eta > 0, \quad \forall 0 \neq \eta \in \mathbb{R}^n.$$

由定理11.5.2(取$X = \varepsilon I$)知存在$\lambda_0 > 0$使得$\varepsilon \lambda_0^2 I + \lambda_0 Y + Z \leqslant 0$, 从而$\lambda_0 Y + Z < 0$. □

11.6 含范数有界不确定性的线性矩阵不等式

在鲁棒控制中称满足$F^T F \leqslant I$的参数矩阵为范数有界不确定性.

引理 11.6.1 设$x \in \mathbb{R}^p$, $y \in \mathbb{R}^q$, D和E是适当维数的实矩阵, 则对任意满足$F^T F \leqslant I$的适当维数矩阵F有

$$2x^T DFEy \leqslant \varepsilon x^T DD^T x + \varepsilon^{-1} y^T E^T Ey, \ \forall \varepsilon > 0.$$

证明 由

$$\begin{aligned} 0 &\leqslant (\varepsilon^{\frac{1}{2}} D^T x - \varepsilon^{-\frac{1}{2}} FEy)^T (\varepsilon^{\frac{1}{2}} D^T x - \varepsilon^{-\frac{1}{2}} FEy) \\ &= \varepsilon x^T DD^T x - 2x^T DFEy + \varepsilon^{-1} y^T E^T F^T FEy \\ &\leqslant \varepsilon x^T DD^T x - 2x^T DFEy + \varepsilon^{-1} y^T E^T Ey \end{aligned}$$

得证. □

引理 11.6.2　对任意给定的向量$x \in \mathbb{R}^p$和$y \in \mathbb{R}^q$有

$$\max\{(x^T F y)^2 : F \in \mathbb{R}^{p\times q}, F^T F \leqslant I\} = (x^T x)(y^T y). \tag{11.6.1}$$

证明　设$F \in \mathbb{R}^{p\times q}$是任意给定的满足$F^T F \leqslant I$的矩阵, 根据Schwarz不等式有

$$|x^T F y| \leqslant \sqrt{(x^T x)(y^T F^T F y)},$$

因此

$$(x^T F y)^2 \leqslant (x^T x)(y^T F^T F y) \leqslant (x^T x)(y^T y).$$

另一方面, 取

$$\widetilde{F} = \frac{x y^T}{\sqrt{x^T x}\sqrt{y^T y}},$$

则$\tilde{F}^T \tilde{F} = I$且

$$x^T \widetilde{F} y = \frac{x^T x y^T y}{\sqrt{x^T x}\sqrt{y^T y}} = \sqrt{x^T x}\sqrt{y^T y},$$

即$(x^T F y)^2$在$\widetilde{F}$处达到上界$(x^T x)(y^T y)$, 故等式(11.6.1)成立. □

从引理11.6.2的证明易得下面的推论:

推论 11.6.1　对任意给定的向量$x \in \mathbb{R}^p$和$y \in \mathbb{R}^q$有

$$\max\{(x^T F y)^2 : F \in \mathbb{R}^{p\times q}, F^T F = I\} = (x^T x)(y^T y).$$

定理 11.6.1　设Y, D和E是给定的有适当维数的矩阵, $Y^T = Y$, 则

$$Y + DFE + E^T F^T D^T < 0$$

对所有满足$F^T F \leqslant I$的矩阵F成立的充要条件是存在一个实数$\varepsilon > 0$使得

$$Y + \varepsilon D D^T + \varepsilon^{-1} E^T E < 0.$$

证明　充分性由引理11.6.1推出, 下证必要性.

由已知条件得

$$\xi^T Y \xi + 2\xi^T D F E \xi < 0,\ \forall \xi \neq 0,$$

使用F的任意性及引理11.6.2推出

$$\xi^T Y \xi + 2\sqrt{\xi^T D D^T \xi}\sqrt{\xi^T E^T E \xi} < 0,\ \forall \xi \neq 0,$$

于是

$$\xi^T Y \xi < -2\sqrt{\xi^T D D^T \xi}\sqrt{\xi^T E^T E \xi},\ \forall \xi \neq 0,$$

从而

$$(\xi^T Y \xi)^2 > 4(\xi^T D D^T \xi)(\xi^T E^T E \xi),\ \forall \xi \neq 0.$$

应用定理11.5.3得证. □

11.7 含线性分式不确定性的线性矩阵不等式

设J是满足$I-JJ^T>0$的已知矩阵, F是满足$F^TF\le I$的未知参数矩阵. 鲁棒控制中称$\Delta:=(I-FJ)^{-1}F$为线性分式不确定性. 易见, 当$J=0$时, 线性分式不确定性退化为范数有界不确定性.

引理 11.7.1　设$\Gamma=\{F(I-JF)^{-1}:F^TF\leqslant I\}$, J满足$I-JJ^T>0$, 则

$$\Gamma=\{J^T(I-JJ^T)^{-1}+\Pi^T(I-JJ^T)^{-\frac{1}{2}}:\Pi^T\Pi\leqslant(I-J^TJ)^{-1}\}.$$

证明　令$\Delta=(I-FJ)^{-1}F$. 由$(I+\Delta J)(I-FJ)=I$得矩阵$I+\Delta J$可逆, 于是$F=(I+\Delta J)^{-1}\Delta$. 由$F^TF\leqslant I$ 得

$$\Delta^T(I+\Delta J)^{-T}(I+\Delta J)^{-1}\Delta\leqslant I,$$

于是

$$\Delta^T(I+\Delta J+J^T\Delta^T+\Delta JJ^T\Delta^T)^{-1}\Delta\leqslant I.$$

使用Schur补引理得到

$$\Delta\Delta^T\leqslant I+\Delta J+J^T\Delta^T+\Delta JJ^T\Delta^T,$$

即

$$\Pi_\Delta^T\Pi_\Delta\leqslant I+J^T(I-JJ^T)^{-1}J=(I-J^TJ)^{-1},$$

其中

$$\Pi_\Delta=(I-JJ^T)^{\frac{1}{2}}\Delta^T-(I-JJ^T)^{-\frac{1}{2}}J.$$

从而引理得证. □

引理 11.7.2　[20] 设M,S,N是给定的有合适维数的矩阵, $M=M^T$, Γ和J同前面定义, 则

$$M+S\Delta N+N^T\Delta^TS^T<0,\quad\forall\Delta\in\Gamma \tag{11.7.1}$$

的充要条件是存在$\varepsilon>0$使得

$$M+\begin{bmatrix}\varepsilon^{-1}N^T & \varepsilon S\end{bmatrix}\begin{bmatrix}I & -J\\ -J^T & I\end{bmatrix}^{-1}\begin{bmatrix}\varepsilon^{-1}N\\ \varepsilon S^T\end{bmatrix}<0. \tag{11.7.2}$$

证明 由引理11.7.1, 矩阵不等式(11.7.1)对所有$\Delta \in \Gamma$成立等价于

$$
\begin{gathered}
M + SJ^T(I - JJ^T)^{-1}N + S\Pi^T(I - JJ^T)^{-\frac{1}{2}}N \\
+N^T(I - JJ^T)^{-1}JS^T + N^T(I - JJ^T)^{-\frac{1}{2}}\Pi S^T < 0
\end{gathered}
$$

对所有满足

$$
\Pi^T\Pi \leqslant (I - J^TJ)^{-1}
$$

的Π成立. 由定理11.6.1知矩阵不等式(11.7.1)成立的充要条件是存在$\varepsilon > 0$使得

$$
\begin{aligned}
&M + SJ^T(I - JJ^T)^{-1}N + N^T(I - JJ^T)^{-1}JS^T \\
&+\varepsilon^{-2}N^T(I - JJ^T)^{-1}N + \varepsilon^2 S(I - J^TJ)^{-1}S^T < 0,
\end{aligned} \tag{11.7.3}
$$

即式(11.7.2)成立. □

定理 11.7.1 设M, S, N为有合适维数的实矩阵, $M = M^T$, Γ和J同前面定义, 则

$$
M + S\Delta N + N^T\Delta^T S^T < 0, \quad \forall \Delta \in \Gamma. \tag{11.7.4}
$$

的充要条件是存在$\delta > 0$使得

$$
\begin{bmatrix}
\delta M & S & \delta N^T \\
S^T & -I & J^T \\
\delta N & J & -I
\end{bmatrix} < 0.
$$

证明 组合Schur补引理和引理11.7.2易得.(留给读者) □

11.8 Jensen不等式

本节介绍Jensen不等式及其扩展版本, 并通过比较时滞控制系统分析和综合中常用的两个不等式说明Jensen不等式的优点.

11.8.1 Jensen不等式

定理 11.8.1 **(连续Jensen不等式)** [21] 设$a, b \in \mathbb{R}$, $b > a$, $M \in \mathbb{R}^{n\times n}$, $M^T = M > 0$, 向量函数$\omega : [a, b] \to \mathbb{R}^n$在区间$[a, b]$上可积, 则

$$
(b - a)\int_a^b \omega^T(s)M\omega(s)\mathrm{d}s \geqslant \left(\int_a^b \omega^T(s)\mathrm{d}s\right) M \left(\int_a^b \omega(s)\mathrm{d}s\right).
$$

证明 由$M>0$及Schur补引理知

$$\begin{bmatrix} \omega^T(\beta)M\omega(\beta) & \omega^T(\beta) \\ \omega(\beta) & M^{-1} \end{bmatrix} \geqslant 0,\ \forall \beta \in [a,b],$$

于是

$$\begin{bmatrix} \int_a^b \omega^T(s)M\omega(s)\mathrm{d}s & \int_a^b \omega^T(s)\mathrm{d}s \\ \int_a^b \omega(s)\mathrm{d}s & \int_a^b M^{-1}\mathrm{d}s \end{bmatrix} \geqslant 0.$$

再由$M>0$及Schur补引理知, 结论得证. □

类似于上面的定理, 可得到下面的离散Jensen不等式.

定理 11.8.2 (**离散Jensen不等式**) [22] 设$Z\in\mathbb{R}^{n\times n}$满足$Z=Z^T>0$, $x_i\in\mathbb{R}^n$, $i=r_1,\cdots,r_2$, r_1和r_2是整数, 则

$$(r_2-r_1+1)\sum_{i=r_1}^{r_2} x_i^T Z x_i \geqslant \left(\sum_{i=r_1}^{r_2} x_i^T\right) Z \left(\sum_{i=r_1}^{r_2} x_i\right).$$

下面给出Jensen不等式的几个扩展版本.

定理 11.8.3 设$E,M\in\mathbb{R}^{n\times n}$满足$M^T=M>0$, $a,b\in\mathbb{R}$, $b>a$, 向量函数$x:[a,b]\to\mathbb{R}^n$在区间$[a,b]$上可导, 则

$$(b-a)\int_a^b \dot{x}^T(s)E^TME\dot{x}(s)\,\mathrm{d}s \geqslant \begin{bmatrix} x(b) \\ x(a) \end{bmatrix}^T \begin{bmatrix} E^TME & -E^TME \\ -E^TME & E^TME \end{bmatrix} \begin{bmatrix} x(b) \\ x(a) \end{bmatrix}. \tag{11.8.1}$$

证明 在连续Jensen不等式中取$\omega(s)=E\dot{x}(s)$即可. □

推论 11.8.1 [23] 设$M\in\mathbb{R}^{n\times n}$满足$M^T=M>0$, $a,b\in\mathbb{R}$, $b>a$, 向量函数$x:[a,b]\to\mathbb{R}^n$在区间$[a,b]$上可导, 则

$$(b-a)\int_a^b \dot{x}^T(s)M\dot{x}(s)\,\mathrm{d}s \geqslant \begin{bmatrix} x(b) \\ x(a) \end{bmatrix}^T \begin{bmatrix} M & -M \\ -M & M \end{bmatrix} \begin{bmatrix} x(b) \\ x(a) \end{bmatrix}.$$

证明 在定理11.8.3中取$E=I_n$得到. □

定理 11.8.4 设$E,W_1,W_2,W_3\in\mathbb{R}^{n\times n}$满足

$$\begin{bmatrix} W_1 & W_2 \\ W_2^T & W_3 \end{bmatrix}^T = \begin{bmatrix} W_1 & W_2 \\ W_2^T & W_3 \end{bmatrix} > 0,$$

$a, b \in \mathbb{R}$, $b > a$, 向量函数$x : [a, b] \to \mathbb{R}^n$在区间$[a, b]$上可导, 则

$$
\begin{aligned}
&(b-a)\int_a^b \left(\begin{bmatrix} x(s) \\ E\dot{x}(s) \end{bmatrix}^T \begin{bmatrix} W_1 & W_2 \\ W_2^T & W_3 \end{bmatrix} \begin{bmatrix} x(s) \\ E\dot{x}(s) \end{bmatrix} \right) \mathrm{d}s \\
\geqslant\ & \begin{bmatrix} x(b) \\ x(a) \\ \int_a^b x(s)\mathrm{d}s \end{bmatrix}^T \begin{bmatrix} E^TW_3E & -E^TW_3E & E^TW_2^T \\ -E^TW_3E & E^TW_3E & -E^TW_2^T \\ W_2E & -W_2E & W_1 \end{bmatrix} \begin{bmatrix} x(b) \\ x(a) \\ \int_a^b x(s)\mathrm{d}s \end{bmatrix}.
\end{aligned}
\tag{11.8.2}
$$

证明 对式(11.8.2)的左端应用连续Jensen不等式, 然后再经恒等变形可得结论. □

定理 11.8.5 [21] 设$M \in \mathbb{R}^{n\times n}$满足$M > 0$, $0 < \alpha < 1$, 向量函数$\omega_1 : [0, \alpha] \to \mathbb{R}^n$在区间$[0, \alpha]$上可积, 向量函数$\omega_2 : [\alpha, 1] \to \mathbb{R}^n$在区间$[\alpha, 1]$上可积, 则

$$
\begin{aligned}
&\textstyle\int_0^1 [(1-\alpha)\omega_1^T(\alpha)M\omega_1(\alpha) + \alpha\omega_2^T(\alpha)M\omega_2(\alpha)]\mathrm{d}\alpha \\
\geqslant\ & \textstyle\int_0^1 \left(\int_0^\alpha \omega_1(\beta)\mathrm{d}\beta + \int_\alpha^1 \omega_2(\beta)\mathrm{d}\beta\right)^T M \left(\int_0^\alpha \omega_1(\beta)\mathrm{d}\beta + \int_\alpha^1 \omega_2(\beta)\mathrm{d}\beta\right) \mathrm{d}\alpha;
\end{aligned}
\tag{11.8.3}
$$

$$
\int_0^1 (1-\alpha)\omega_1^T(\alpha)M\omega_1(\alpha)\mathrm{d}\alpha \geqslant \int_0^1 \left(\int_0^\alpha \omega_1(\beta)\mathrm{d}\beta\right)^T M \left(\int_0^\alpha \omega_1(\beta)\mathrm{d}\beta\right) \mathrm{d}\alpha;
\tag{11.8.4}
$$

$$
\int_0^1 \alpha\omega_2^T(\alpha)M\omega_2(\alpha)\mathrm{d}\alpha \geqslant \int_0^1 \left(\int_\alpha^1 \omega_2(\beta)\mathrm{d}\beta\right)^T M \left(\int_\alpha^1 \omega_2(\beta)\mathrm{d}\beta\right) \mathrm{d}\alpha.
\tag{11.8.5}
$$

证明 令

$$
\omega(\beta) = \begin{cases} \omega_1(\beta), & \text{若} 0 \leqslant \beta \leqslant \alpha, \\ \omega_2(\beta), & \text{若} \alpha < \beta \leqslant 1. \end{cases}
$$

应用定理11.8.1得到

$$
\begin{aligned}
&\textstyle\int_0^\alpha \omega_1^T(\beta)M\omega_1(\beta)\mathrm{d}\beta + \int_\alpha^1 \omega_2^T(\beta)M\omega_2(\beta)\mathrm{d}\beta \\
\geqslant\ & \textstyle\left(\int_0^\alpha \omega_1(\beta)\mathrm{d}\beta + \int_\alpha^1 \omega_2(\beta)\mathrm{d}\beta\right)^T M \left(\int_0^\alpha \omega_1(\beta)\mathrm{d}\beta + \int_\alpha^1 \omega_2(\beta)\mathrm{d}\beta\right).
\end{aligned}
$$

将上式两边对α从0到1积分并交换积分次序得式(11.8.3). 在式(11.8.3)中取$\omega_2 = 0$得式(11.8.4)、取$\omega_1 = 0$得式(11.8.5). □

读者可以沿着前面的讨论给出离散Jensen不等式的扩展版本.

11.8.2 两个不等式的比较

定理 11.8.6 设$E, M_1, M_2, X \in \mathbb{R}^{n\times n}$, $Z \in \mathbb{R}^{2n\times 2n}$, $a, b \in \mathbb{R}$, $a < b$, $x : [a, b] \to \mathbb{R}^n$是可导的向量函数. 若

$$\begin{bmatrix} X & Y \\ Y^T & Z \end{bmatrix} \geqslant 0,\ Y = \begin{bmatrix} M_1 & M_2 \end{bmatrix}, \tag{11.8.6}$$

则

$$-\int_a^b \dot{x}^T(s)E^T XE\dot{x}(s)\mathrm{d}s \leqslant \xi^T(t)\Gamma\xi(t) + (b-a)\xi^T(t)Z\xi(t), \tag{11.8.7}$$

其中

$$\Gamma = \begin{bmatrix} M_1^T E + E^T M_1 & -M_1^T E + E^T M_2 \\ -E^T M_1 + M_2^T E & -M_2^T E - E^T M_2 \end{bmatrix},\quad \xi(t) = \begin{bmatrix} x(b) \\ x(a) \end{bmatrix}. \tag{11.8.8}$$

证明 由式(11.8.6)得

$$\begin{bmatrix} E\dot{x}(s) \\ \xi(t) \end{bmatrix}^T \begin{bmatrix} X & Y \\ Y^T & Z \end{bmatrix} \begin{bmatrix} E\dot{x}(s) \\ \xi(t) \end{bmatrix} \geqslant 0,\quad \forall s \in [a, b],$$

于是

$$0 \leqslant \dot{x}^T(s)E^T XE\dot{x}(s) + 2\xi^T(t)Y^T E\dot{x}(s) + \xi^T(t)Z\xi(t),\quad \forall s \in [a, b].$$

两边积分得

$$0 \leqslant \int_a^b \dot{x}^T(s)E^T XE\dot{x}(s)\mathrm{d}s + 2\int_a^b \xi^T(t)Y^T E\dot{x}(s)\mathrm{d}s + \int_a^b \xi^T(t)Z\xi(t)\mathrm{d}s, \tag{11.8.9}$$

从而

$$-\int_a^b \dot{x}^T(s)E^T XE\dot{x}(s)\mathrm{d}s \leqslant 2\xi^T(t)Y^T E(x(b) - x(a)) + (b-a)\xi^T(t)Z\xi(t), \tag{11.8.10}$$

即

$$-\int_a^b \dot{x}^T(s)E^T XE\dot{x}(s)\mathrm{d}s \leqslant (b-a)\xi^T(t)Z\xi(t) + 2\xi^T(t)Y^T \begin{bmatrix} E & -E \end{bmatrix} \xi(t). \tag{11.8.11}$$

在式(11.8.11)中取$Y = \begin{bmatrix} M_1 & M_2 \end{bmatrix}$得证. □

积分不等式(11.8.7)在不等式约束(11.8.6)下成立, 当$X > 0$时约束条件可以被替换. 例如: 取$Z = Y^T X^{-1} Y$, 则得下面的推论.

推论 11.8.2 [24] 设$E, M_1, M_2, M \in \mathbb{R}^{n\times n}$, $M^T = M > 0$, $a, b \in \mathbb{R}$, $a < b$, $x : [a, b] \to \mathbb{R}^n$是可导的向量函数, 则

$$-\int_a^b \dot{x}^T(s)E^TME\dot{x}(s)\mathrm{d}s \leqslant \xi^T(t)\Gamma\xi(t) + (b-a)\xi^T(t)Y^TM^{-1}Y\xi(t), \tag{11.8.12}$$

其中Γ, $\xi(t)$, Y同定理11.8.6中定义.

注记 11.8.1 注意到$\Gamma = Y^T\begin{bmatrix} E & -E\end{bmatrix} + \begin{bmatrix} E & -E\end{bmatrix}^T Y$, 易见

$$\begin{aligned} & \xi^T(t)\Gamma\xi(t) + (b-a)\xi^T(t)Y^TM^{-1}Y\xi(t) \\ \geqslant\ & -(b-a)^{-1}\xi^T(t)[E \quad -E]^TM[E \quad -E]\xi(t) \\ =\ & -(b-a)^{-1}\begin{bmatrix} x(b) \\ x(a)\end{bmatrix}^T \begin{bmatrix} E^TME & -E^TME \\ -E^TME & E^TME\end{bmatrix}\begin{bmatrix} x(b) \\ x(a)\end{bmatrix}, \end{aligned}$$

故不等式(11.8.1)比不等式(11.8.12)给出了积分$-\int_a^b \dot{x}^T(s)E^TME\dot{x}(s)\mathrm{d}s$的更加精确的估计. 因而, 应用式(11.8.1)处理时滞系统分析和综合中一些问题可能会得到更好的结果.

习 题

1. 设Q和R均为实对称矩阵, 则$\begin{bmatrix} Q & S \\ S^T & R\end{bmatrix} \geqslant 0$的充要条件是$R \geqslant 0$, $Q - SR^+S^T \geqslant 0$且$S(I - RR^+) = 0$.

2. 设P_{ii} $(i = 1, 2, 3)$是实对称矩阵, P_{ij} $(1 \leqslant i < j \leqslant 3)$是实矩阵, 并且这些矩阵有合适的维数, 则存在实对称矩阵X使得

$$\begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{12}^T & P_{22} + X & P_{23} \\ P_{13}^T & P_{23}^T & P_{33}\end{bmatrix} < 0$$

有解的充要条件是

$$\begin{bmatrix} P_{11} & P_{13} \\ P_{13}^T & P_{33}\end{bmatrix} < 0.$$

3. 设$A \in \mathbb{R}^n$, $B \in \mathbb{R}^{n\times r}$, $C \in \mathbb{R}^{m\times n}$, γ为给定的正数, 则存在$S > 0$使得

$$\begin{bmatrix} A^TS + SA & SB & C^T \\ (SB)^T & -\gamma I & 0 \\ C & 0 & -\gamma I\end{bmatrix} < 0 \tag{11.8.13}$$

的充要条件是存在矩阵F, G和$S>0$使得

$$\begin{bmatrix} A^TG+G^TA & \Phi & G^TB & C^T \\ \Phi^T & -F-F^T & F^TB & 0 \\ B^TG & B^TF & -\gamma I & 0 \\ C & 0 & 0 & -\gamma I \end{bmatrix} < 0, \tag{11.8.14}$$

其中$\Phi = S - G^T + A^TF$.

(提示: 使用Projection引理.)

4. 证明: 对任意的$x\in\mathbb{R}^n$和适当维数的矩阵P, D, F和E有

$$\max\{(x^TPDFEx)^2 : F^TF \leqslant I\} = x^TPDD^TP^Txx^TE^TEx. \tag{11.8.15}$$

第12章　代数Riccati矩阵方程

本章介绍两类代数Riccati矩阵方程的实对称(半正定)稳定解的存在性和性质，这些内容是求解连续线性系统的H_2和H_∞优化问题的理论基础.

12.1 Lyapunov矩阵方程

12.1.1 矩阵对的能稳性和能检测性

定义 12.1.1　设$A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$和$C \in \mathbb{R}^{l\times n}$. 若矩阵对$(A, B)$满足

$$\operatorname{rank}[sI - A \quad B] = n,\ \forall s \in \overline{\mathbb{C}^+},$$

称矩阵对(A, B)能稳; 若矩阵对(A, C)满足

$$\operatorname{rank}\begin{bmatrix} sI - A \\ C \end{bmatrix} = n,\ \forall s \in \overline{\mathbb{C}^+}, \tag{12.1.1}$$

称矩阵对(A, C)能检测.

命题 12.1.1　设矩阵A, B有合适的维数, 则(A, B)能稳的充分必要条件是对于任意满足$\mathcal{R}(D) \subseteq \mathcal{R}(B)$的$D$和实对称正定矩阵$U$有

$$(A + D, BUB^T),\ (A + D, BU^{\frac{1}{2}})$$

均能稳.

证明　由矩阵$BU^{\frac{1}{2}}$的奇异值分解易见. □

命题 12.1.2 设矩阵A, C有合适的维数, 则(A,C)能检测的充分必要条件是对于任意满足$\mathcal{R}(D^T)\subseteq\mathcal{R}(C^T)$的$D$和实对称正定矩阵$U$有$(A+D,C^TUC)$和$(E,A+D,U^{\frac{1}{2}}C)$均能检测.

证明 由矩阵$U^{\frac{1}{2}}C$的奇异值分解易见. □

12.1.2 连续Lyapunov矩阵方程

引理 12.1.1 若$A,M\in\mathbb{R}^{n\times n}$, $M\geqslant 0$, (A,M)能检测, Lyapunov矩阵方程

$$A^TX+X^TA+M=0 \tag{12.1.2}$$

有实对称(半)正定解X, 则$\lambda(A)\subset\mathbb{C}^-$.

证明 设λ是A的任意一个特征根, y是其对应的特征向量, 则$Ay=\lambda y$, 于是$y^*A^TXy+y^*XAy+y^*My=0$, 从而$2\mathrm{Re}(\lambda)y^*Xy+y^*My=0$.

情形1: $y^*My=0$. 由$M\geqslant 0$得$My=0$, 因而$\begin{bmatrix}\lambda I-A\\ M\end{bmatrix}y=0$. 由$(A,M)$能检测和$y\neq 0$推出$\mathrm{Re}(\lambda)<0$, 从而$\lambda(A)\subset\mathbb{C}^-$.

情形2: $y^*My>0$. 显然$\mathrm{Re}(\lambda)y^*Xy<0$. 使用$X\geqslant 0$得$\mathrm{Re}(\lambda)<0$, 从而$\lambda(A)\subset\mathbb{C}^-$. □

下面的推论由引理12.1.1的证明得到.

推论 12.1.1 若A, $M\in\mathbb{R}^{n\times n}$, $M>0$, Lyapunov矩阵方程(12.1.2)有实对称(半)正定解X, 则$\lambda(A)\subset\mathbb{C}^-$.

定理 12.1.1 若A, $M\in\mathbb{R}^{n\times n}$, A稳定(即A的所有特征根的实部小于零), $M^T=M$, 则$\displaystyle\int_0^\infty \mathrm{e}^{A^Tt}M\mathrm{e}^{At}\mathrm{d}t$是矩阵方程(12.1.2)的唯一实对称解.

证明 由A稳定知

$$\begin{aligned}
&A^T\int_0^\infty \mathrm{e}^{A^Tt}M\mathrm{e}^{At}\mathrm{d}t+\int_0^\infty \mathrm{e}^{A^Tt}M\mathrm{e}^{At}\mathrm{d}tA\\
=&\int_0^\infty (A^T\mathrm{e}^{A^Tt}M\mathrm{e}^{At}+\mathrm{e}^{A^Tt}M\mathrm{e}^{At}A)\mathrm{d}t\\
=&\int_0^\infty \mathrm{d}(\mathrm{e}^{A^Tt}M\mathrm{e}^{At})\\
=&-M,
\end{aligned}$$

即$\int_0^{\infty} e^{A^Tt}Me^{At}dt$是Lyapunov矩阵方程(12.1.2)的实对称解.

下证解的唯一性. 令$\overline{X}$是Lyapunov矩阵方程(12.1.2)的任意实对称解, 则

$$\begin{aligned}\int_0^{\infty} e^{A^Tt}Me^{At}dt &= -\int_0^{\infty} e^{A^Tt}(A^T\overline{X}+\overline{X}A)e^{At}dt\\ &= -\int_0^{\infty} d(e^{A^Tt}\overline{X}e^{At})\\ &= \overline{X},\end{aligned}$$

这说明Lyapunov矩阵方程(12.1.2)的实对称解唯一. □

定理 12.1.2 若$A\in\mathbb{R}^{n\times n}$稳定, $M^T=M\in\mathbb{R}^{n\times n}$, 则

(i) 对于$M>0(M\geqslant 0)$, $\int_0^{\infty} e^{A^Tt}Me^{At}dt$是Lyapunov矩阵方程(12.1.2)的唯一实对称(半)正定解;

(ii) 对于$M<0(M\leqslant 0)$, $\int_0^{\infty} e^{A^Tt}Me^{At}dt$是Lyapunov矩阵方程(12.1.2)的唯一实对称(半)负定解.

证明 下面只证明$M\geqslant 0$的情形, 其他情形的证明类似.

由定理12.1.1, Lyapunov矩阵方程(12.1.2)有唯一实对称解

$$\int_0^{\infty} e^{A^Tt}Me^{At}dt.$$

又因为M是半正定的, 存在矩阵N使得$M=N^TN$. 对于任意的n维向量y, 易见

$$\begin{aligned}y^T\left(\int_0^{\infty} e^{A^Tt}Me^{At}dt\right)y &= \int_0^{\infty}(y^Te^{A^Tt}N^T)(Ne^{At}y)dt\\ &= \int_0^{\infty}\|Ne^{At}y\|_2^2dt,\end{aligned}$$

由$\|Ne^{At}y\|_2^2\geqslant 0$得$\int_0^{\infty} e^{A^Tt}Me^{At}dt$是半正定的. □

注记 12.1.1 当$A\in\mathbb{C}^{n\times n}$, $M^*=M\in\mathbb{C}^{n\times n}$时, 代数Lyapunov方程$A^*X+X^*A+M=0$与式(12.1.2)有很多类似的性质, 故有与上面类似的结论.

12.2 Hamilton矩阵

记

$$\mathcal{J} = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}.$$

定义 12.2.1 若$H \in \mathbb{R}^{2n\times 2n}$满足$\mathcal{J}^{-1}H^T\mathcal{J} = -H$, 称$H$为Hamilton矩阵.

从Hamilton矩阵的定义易见下面的两个命题成立.

命题 12.2.1 设$H \in \mathbb{R}^{2n\times 2n}$是Hamilton矩阵, $\lambda \in \mathbb{C}$. 若λ是H的特征根, 则$-\lambda$也是H的特征根.

命题 12.2.2 若$H \in \mathbb{R}^{2n\times 2n}$是Hamilton矩阵, 则存在矩阵$A \in \mathbb{R}^{n\times n}$和对称矩阵$R, M \in \mathbb{R}^{n\times n}$使得

$$H = \begin{bmatrix} A & R \\ M & -A^T \end{bmatrix}.$$

定义 12.2.2 设$\text{diag}(J_1, J_2)$是方阵A的Jordan标准形. 如果J_1的对角元实部均为负且J_2的对角元实部均非负, 称J_1是矩阵A的**Jordan标准形的稳定部分**.

若Hamilton矩阵$H \in \mathbb{R}^{2n\times 2n}$在虚轴上无特征根, 则由命题12.2.1知$H$的所有特征根可表示为

$$\lambda_1,\ \cdots,\ \lambda_n,\ -\lambda_1,\ \cdots,\ -\lambda_n,$$

其中$\text{Re}(\lambda_i) < 0$, $i = 1, 2, \cdots, n$, 于是H的Jordan标准形可表示为$\text{diag}(J, \hat{J})$, 其中J的对角元素为$\lambda_1, \cdots, \lambda_n$(即$J$是稳定部分); $\hat{J}$的对角元素为$-\lambda_1, \cdots, -\lambda_n$. 再由Jordan标准形理论得存在可逆阵$T$使得

$$H = T\text{diag}(J, \hat{J})T^{-1}.$$

将T写做$\begin{bmatrix} T_1 & T_2 \end{bmatrix}$, 其中$T_1, T_2 \in \mathbb{R}^{2n\times n}$, 由定理2.5.2知$\mathcal{R}(T_1)$和$\mathcal{R}(T_2)$均由矩阵$H$唯一确定, 分别记作$\mathcal{X}_-(H)$和$\mathcal{X}_+(H)$, 称$\mathcal{X}_-(H)$为Hamilton矩阵$H$的稳定子空间. 定理2.5.2也表明$\dim \mathcal{X}_-(H) = \dim \mathcal{X}_+(H) = n$且

$$\mathbb{C}^{2n} = \mathcal{X}_-(H) \oplus \mathcal{X}_+(H).$$

取$\mathcal{X}_-(H)$的一组基, 并将其排成矩阵$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, 其中$X_1, X_2 \in \mathbb{C}^{n\times n}$, 则

$$\mathcal{X}_-(H) = \mathcal{R}\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right).$$

若H满足

$$\mathbb{C}^{2n} = \mathcal{X}_-(H) \oplus \mathcal{R}\left(\begin{bmatrix} 0 \\ I_n \end{bmatrix}\right), \tag{12.2.1}$$

则X_1可逆, 并且$X_2X_1^{-1}$由H唯一确定(即$X_2X_1^{-1}$与$\mathcal{X}_-(H)$的基的选取无关), 将其记作$\mathrm{Ric}(H)$. 用$\mathrm{dom}(\mathrm{Ric})$表示在虚轴上无特征值且满足式(12.2.1)的所有Hamilton矩阵的集合. 将在12.3节中展示: 若$H \in \mathrm{dom}(\mathrm{Ric})$, 则$\mathrm{Ric}(H)$是实对称阵, 并且$A + R\mathrm{Ric}(H)$是稳定的.

对于$A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$和$C \in \mathbb{R}^{q\times n}$, 记

$$H_1 = \begin{bmatrix} A & -BB^T \\ -C^TC & -A^T \end{bmatrix},$$

则H_1是Hamilton矩阵.

定理 12.2.1 λ是H_1在虚轴上的特征根的充分必要条件是λ是(A, B)在虚轴上的不能控特征根或λ是(A, C)在虚轴上的不能观特征根.

证明 **必要性**. 设$j\omega$是H_1的一个特征根, 且$\begin{bmatrix} x \\ y \end{bmatrix} \neq 0$是相应的特征向量, 则

$$\begin{bmatrix} A & -BB^T \\ -C^TC & -A^T \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = j\omega\begin{bmatrix} x \\ y \end{bmatrix},$$

于是

$$Ax - BB^Ty = j\omega x, \tag{12.2.2a}$$

$$-C^TCx - A^Ty = j\omega y. \tag{12.2.2b}$$

对式(12.2.2a)和式(12.2.2b)分别左乘y^*和x^*, 然后相加得

$$-(x^*C^TCx + y^*BB^Ty) = [x^*A^Ty - (x^*A^Ty)^*] + j\omega[x^*y + (x^*y)^*].$$

注意到上式左端的两项是实数, 但右端却是零或纯虚数, 所以

$$x^*C^TCx + y^*BB^Ty = 0,$$

因此

$$y^*B = 0,\ Cx = 0.$$

进而由式(12.2.2)得

$$Ax = j\omega x,\ y^*A = j\omega y^*.$$

当$y \neq 0$时, $j\omega$为(A,B)的不能控特征根; 当$y=0$时, 由$\begin{bmatrix} x \\ y \end{bmatrix} \neq 0$推出$x \neq 0$, 从而$j\omega$为$(A,C)$的不能观特征根.

充分性. 由必要性的逆过程得到. □

推论 12.2.1 若(A,B)能稳且(A,C)能检测, 则H_1在虚轴上没有特征根.

12.3 代数Riccati矩阵方程的实对称稳定解

定义 12.3.1 对于A, R, $M \in \mathbb{R}^{n\times n}$满足$R=R^T$及$M=M^T$, 称方程

$$A^TX + X^TA + X^TRX + M = 0 \tag{12.3.1}$$

为代数Riccati矩阵方程.

定义 12.3.2 称X是代数Riccati方程(12.3.1)的一个稳定解, 若$X \in \mathbb{R}^{n\times n}$满足式(12.3.1)且$A+RX$稳定.

记

$$H = \begin{bmatrix} A & R \\ -M & -A^T \end{bmatrix}.$$

显然H是Hamilton矩阵.

定理 12.3.1 设$H \in \mathrm{dom}(\mathrm{Ric})$, $X = \mathrm{Ric}(H)$. 证明:

(i) 存在$T_2 \in \mathbb{C}^{n\times n}$和可逆的$T_1 \in \mathbb{C}^{n\times n}$满足

$$X = T_2T_1^{-1},\ H\begin{bmatrix} T_1 \\ T_2 \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} J,$$

其中J是H的Jordan标准形的稳定部分;

(ii) X是代数Riccati矩阵方程(12.3.1)的实对称稳定解;

(iii) $\mathcal{X}_-(H) = \mathcal{R}\left(\begin{bmatrix} I_n \\ X \end{bmatrix}\right)$.

证明 (i) 由Ric(H)的定义易见(见12.2节).

(ii) 由(i)得

$$AT_1 + RT_2 = T_1 J, \tag{12.3.2}$$

$$-MT_1 - A^T T_2 = T_2 J. \tag{12.3.3}$$

在式(12.3.2)两端左乘T_2^T、式(12.3.3)两端左乘T_1^T得

$$T_2^T AT_1 + T_2^T RT_2 = T_2^T T_1 J,$$

$$-T_1^T MT_1 - T_1^T A^T T_2 = T_1^T T_2 J,$$

于是

$$(T_2^T T_1 - T_1^T T_2)J = T_2^T AT_1 + T_1^T A^T T_2 + T_2^T RT_2 + T_1^T MT_1. \tag{12.3.4}$$

因为上式右端是对称的, 则左端也是对称的, 于是

$$(T_2^T T_1 - T_1^T T_2)J = J^T (T_2^T T_1 - T_1^T T_2)^T = -J^T (T_2^T T_1 - T_1^T T_2),$$

即$(T_2^T T_1 - T_1^T T_2)J + J^T (T_2^T T_1 - T_1^T T_2) = 0$. 由$J$稳定及定理12.1.2得

$$T_2^T T_1 = T_1^T T_2. \tag{12.3.5}$$

同理

$$T_2^* T_1 = T_1^* T_2. \tag{12.3.6}$$

由式(12.3.5)得$(T_2 T_1^{-1})^T = T_1^{-T} T_2^T = T_2 T_1^{-1}$, 故$T_2 T_1^{-1}$是对称的. 这和式(12.3.6)一起推出

$$\overline{T_2 T_1^{-1}} = \overline{T_2}\,\overline{T_1^{-1}} = (T_2^T)^* (T_1^{-T})^* = ((T_2 T_1^{-1})^T)^* = (T_2 T_1^{-1})^* = T_2 T_1^{-1},$$

即$T_2 T_1^{-1}$是实的, 从而$T_2 T_1^{-1}$是实对称的.

组合式(12.3.4)和式(12.3.5)得到

$$T_2^T AT_1 + T_1^T A^T T_2 + T_2^T RT_2 + T_1^T MT_1 = 0.$$

左乘T_1^{-T}并且右乘T_1^{-1}得到

$$(T_2 T_1^{-1})^T A + A^T (T_2 T_1^{-1}) + (T_2 T_1^{-1})^T R (T_2 T_1^{-1}) + M = 0,$$

即$T_2 T_1^{-1}$是矩阵方程(12.3.1)的解.

由式(12.3.2)得$A + R(T_2 T_1^{-1}) = (AT_1 + RT_2)T_1^{-1} = T_1 J T_1^{-1}$, 使用$J$稳定推出$A + RT_2 T_1^{-1}$稳定.

总之, $T_2 T_1^{-1}$是代数Riccati矩阵方程(12.3.1)的实对称稳定解.

(iii) 由$\mathcal{X}_-(H)$的定义和(i)易见. □

定理 12.3.2 设矩阵方程(12.3.1)有实对称稳定解, 则矩阵H在虚轴上无特征根且(A,R)能稳.

证明 设P是代数Riccati矩阵方程(12.3.1)的实对称稳定解, 则$A^TP+PA+PRP+M=0$且$A+RP$稳定. 令$T=\begin{bmatrix} I & 0 \\ P & I \end{bmatrix}$, 则

$$\begin{aligned} T^{-1}HT &= \begin{bmatrix} I & 0 \\ -P & I \end{bmatrix}\begin{bmatrix} A & R \\ -M & -A^T \end{bmatrix}\begin{bmatrix} I & 0 \\ P & I \end{bmatrix} \\ &= \begin{bmatrix} A+RP & R \\ 0 & -(A+RP)^T \end{bmatrix}. \end{aligned}$$

由$A+RP$稳定知H在虚轴上无特征根且(A,R)能稳. □

对于$\eta\in\{1,-1\}$和$B\in\mathbb{R}^{n\times m}$, 设$R=-\eta BB^T$, 则式(12.3.1)变成

$$A^TX+X^TA-\eta X^TBB^TX+M=0. \tag{12.3.7}$$

记

$$H_\eta=\begin{bmatrix} A & -\eta BB^T \\ -M & -A^T \end{bmatrix},$$

显然H_η是Hamilton矩阵.

引理 12.3.1 若H_η在虚轴上无特征根, (A,B)能稳, 则$H_\eta\in\mathrm{dom(Ric)}$.

证明 由H_η在虚轴上无特征根及12.2节的讨论知存在矩阵T_1和T_2使得$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix}$列满秩且

$$H_\eta\begin{bmatrix} T_1 \\ T_2 \end{bmatrix}=\begin{bmatrix} T_1 \\ T_2 \end{bmatrix}J,$$

其中J是H_η的Jordan标准形的稳定部分, 于是

$$AT_1-\eta BB^TT_2=T_1J, \tag{12.3.8}$$

$$-MT_1-A^TT_2=T_2J. \tag{12.3.9}$$

在式(12.3.8)两端左乘T_2^*、式(12.3.9)两端左乘T_1^*得

$$T_2^*AT_1-\eta T_2^*BB^TT_2=T_2^*T_1J,$$

$$-T_1^*MT_1-T_1^*A^TT_2=T_1^*T_2J.$$

于是

$$(T_2^*T_1-T_1^*T_2)J=T_2^*AT_1-\eta T_2^*BB^TT_2+T_1^*MT_1+T_1^*A^TT_2. \tag{12.3.10}$$

由右端是Hermite阵知左端也是Hermite阵, 故

$$(T_2^*T_1 - T_1^*T_2)J + J^*(T_2^*T_1 - T_1^*T_2) = 0.$$

使用J稳定、定理12.1.2和注记12.1.1推出

$$T_2^*T_1 = T_1^*T_2. \tag{12.3.11}$$

若T_1不可逆, 则$\mathcal{N}(T_1) \neq \{0\}$. 任取$0 \neq w \in \mathcal{N}(T_1)$, 并在式(12.3.8)两端左乘$w^*T_2^*$、右乘$w$得

$$-\eta w^*T_2^*BB^TT_2w = w^*T_2^*T_1Jw. \tag{12.3.12}$$

在式(12.3.11)两端左乘w^*得$w^*T_2^*T_1 = 0$, 于是由式(12.3.12)得

$$B^TT_2w = 0. \tag{12.3.13}$$

再右乘w于式(12.3.8)两端, 得$T_1Jw = 0$. 因而$\mathcal{N}(T_1)$是J的不变子空间. 从而J可看成是$\mathcal{N}(T_1)$上的线性变换. 故存在$\lambda_0 \in \mathbb{C}$和$0 \neq z \in \mathcal{N}(T_1)$使得$Jz = \lambda_0 z$. 由$J$稳定知$\mathrm{Re}(\lambda_0) < 0$. 右乘$z$于式(12.3.9)两端得

$$-A^TT_2z = T_2Jz = \lambda_0T_2z,$$

于是$(-\lambda_0I - A^T)T_2z = 0$, 从而

$$z^*T_2^*(-\overline{\lambda_0}I - A) = 0. \tag{12.3.14}$$

在式(12.3.13)中选择$w = z$得$z^*T_2^*B = 0$, 这和式(12.3.14)一起推出

$$z^*T_2^*\begin{bmatrix} -\overline{\lambda_0}I - A & B \end{bmatrix} = 0.$$

使用(A, B)能稳得到$z^*T_2^* = 0$, 即$T_2z = 0$. 这和$z \in \mathcal{N}(T_1)$一起推出$z \in \mathcal{N}\left(\begin{bmatrix} T_1 \\ T_2 \end{bmatrix}\right)$. 由$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix}$列满秩得$z = 0$, 矛盾. 故$T_1$可逆, 从而$H_\eta \in \mathrm{dom(Ric)}$. □

定理 12.3.3 设H_η在虚轴上无特征根, (A, B)能稳, 则$\mathrm{Ric}(H)$是矩阵方程(12.3.7)的唯一的实对称稳定解. 特别地, 若(A, M)能观, M实对称半正(负)定, 则$\mathrm{Ric}(H)$可逆.

证明 组合引理12.3.1和定理12.3.1得到$\mathrm{Ric}(H)$是矩阵方程(12.3.7)的实对称稳定解.

若X_1和X_2是矩阵方程(12.3.7)的两个实对称稳定解, 则

$$A^TX_i + X_iA - \eta X_iBB^TX_i + M = 0, \ i = 1, 2,$$

于是

$$\begin{aligned} &(A - \eta BB^TX_1)^T(X_1 - X_2) + (X_1 - X_2)(A - \eta BB^TX_1) \\ + \ &\eta(X_1 - X_2)BB^T(X_1 - X_2) = 0. \end{aligned}$$

这表明$X_1 - X_2$是矩阵方程

$$(A - \eta BB^T X_1)^T Y + Y(A - \eta BB^T X_1) + \eta(X_1 - X_2)BB^T(X_1 - X_2) = 0$$

的实对称解. 既然$A - \eta BB^T X_1$稳定, 使用$(X_1 - X_2)BB^T(X_1 - X_2) \geqslant 0$和定理12.1.2推出$X_1 \geqslant X_2$. 同理$X_2 \geqslant X_1$. 因而$X_1 = X_2$.

总之, 代数Riccati矩阵方程(12.3.7)有唯一的实对称稳定解.

特别地, 若(A, M)能观, M实对称半正(负)定, 并且$\mathrm{Ric}(H)$不可逆. 不妨设$\mathrm{Ric}(H) = \mathrm{diag}(P, 0)$, 其中$P$可逆. 对应地令

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, \ B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \ M = \begin{bmatrix} M_1 & M_2 \\ M_3 & M_4 \end{bmatrix},$$

则由式(12.3.7)得$A_2 = 0$且$M_4 = 0$, 于是由M实对称半正(负)定推出$M = \mathrm{diag}(M_1, 0)$, 从而

$$\mathrm{rank} \begin{bmatrix} \lambda I - A \\ M \end{bmatrix} = \mathrm{rank} \begin{bmatrix} \lambda I - A_1 & 0 \\ -A_3 & \lambda I - A_4 \\ M_1 & 0 \end{bmatrix}, \ \forall \lambda \in \mathbb{C}.$$

因而对于A_4的特征根λ_0有

$$\mathrm{rank} \begin{bmatrix} \lambda_0 I - A \\ M \end{bmatrix} < n,$$

与(A, M)能观矛盾. 故$\mathrm{Ric}(H)$可逆. □

组合定理12.3.2、命题12.1.1和定理12.3.3得到

推论 12.3.1 H_η在虚轴上无特征根且(A, B)能稳的充要条件是矩阵方程(12.3.7)有实对称稳定解.

12.4 H_2代数Riccati矩阵方程的实对称半正定稳定解

定义 12.4.1 对于$A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$和$C \in \mathbb{R}^{l \times n}$, 称方程

$$A^T X + X^T A - X^T BB^T X + C^T C = 0 \tag{12.4.1}$$

为H_2代数Riccati矩阵方程.

定理 12.4.1 设H_1在虚轴上无特征根, (A, B)能稳, 则H_2代数Riccati矩阵方程(12.4.1)有唯一的实对称半正定稳定解.

证明 由定理12.3.3知H_2代数Riccati矩阵方程(12.4.1)有唯一的实对称稳定解P. 因而只需证$P \geqslant 0$. 事实上, 由$A^TP + PA - PBB^TP + C^TC = 0$得$(A-BB^TP)^TP+P(A-BB^TP)+PBB^TP+C^TC=0$. 使用$PBB^TP+C^TC \geqslant 0$, $A-BB^TP$稳定及定理12.1.2得到$P \geqslant 0$. □

定理 12.4.2 设(A,B)能稳, (A,C)能检测, $\mathcal{X}$表示(A,C)的不能观子空间, 则

(i) $H := \begin{bmatrix} A & -BB^T \\ -C^TC & -A^T \end{bmatrix} \in \mathrm{dom(Ric)}$;

(ii) $X := \mathrm{Ric}(H)$是代数Riccati矩阵方程(12.4.1)的唯一的实对称半正定稳定解;

(iii) $\mathcal{N}(X) \subset \mathcal{X} \subset \mathcal{N}(C)$;

(iv) 存在矩阵M满足$XM = C^T$;

(v) 特别地, 当(A,C)能观时代数Riccati矩阵方程(12.4.1)有唯一的实对称正定稳定解.

证明 (i) 由推论12.2.1知H在虚轴上无特征根. 再由(A,B)能稳和引理12.3.1得证.

(ii) 组合(i)和定理12.4.1得到.

(iii) 由$\mathcal{X} = \mathcal{N}\left(\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}\right)$得到$\mathcal{X} \subset \mathcal{N}(C)$. 下证$\mathcal{N}(X) \subset \mathcal{X}$.

若X可逆, 则$\mathcal{N}(X) \subset \mathcal{X}$显然. 若$X$不可逆, 由$X$实对称半正定不妨设$X = \mathrm{diag}(X_0, 0)$, 其中$X_0 > 0$. 使用(ii)推出

$$C = \begin{bmatrix} C_1 & 0 \end{bmatrix},\ A = \begin{bmatrix} A_1 & 0 \\ A_3 & A_4 \end{bmatrix},$$

于是

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} = \begin{bmatrix} C_1 & 0 \\ C_1A_1 & 0 \\ \vdots & \vdots \\ C_1A_1^{n-1} & 0 \end{bmatrix},$$

故$\mathcal{N}(X) \subset \mathcal{X}$.

(iv) 由(iii)易见$\mathcal{R}(C^T) = \mathcal{N}(C)^\perp \subseteq \mathcal{N}(X)^\perp = \mathcal{R}(X)$, 故存在矩阵$M$满足$XM = C^T$.

(v) 由(ii)和(iii)的证明得到. □

12.5 H_∞范数与H_∞代数Riccati矩阵方程

12.5.1 H_∞范数与H_∞代数Riccati矩阵方程的定义

对于定常线性系统

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{12.5.1a}$$

$$y(t) = Cx(t) + Du(t). \tag{12.5.1b}$$

记其传递函数矩阵$C(sI-A)^{-1}B+D$为$G(s)$. 定义$G(s)$的H_∞范数$\|G(s)\|_\infty$为

$$\|G(s)\|_\infty = \sup_{\omega\in\mathbb{R}} \sigma_{\max}(G(j\omega)),$$

其中$\sigma_{\max}(\cdot)$表示矩阵的最大奇异值.

定义 12.5.1 对于$\gamma > 0$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$和$C \in \mathbb{R}^{l\times n}$, 称方程

$$A^TX + X^TA + \gamma^{-2}X^TBB^TX + C^TC = 0 \tag{12.5.2}$$

为H_∞代数Riccati矩阵方程.

记

$$H_M = \begin{bmatrix} A_M & B_MB_M^T \\ -C_M^TC_M & -A_M^T \end{bmatrix}, \tag{12.5.3}$$

其中

$$\begin{cases} A_M = A + BD_M^{-1}D^TC, \\ B_M = BD_M^{-\frac{1}{2}}, \\ C_M = (I + DD_M^{-1}D^T)^{\frac{1}{2}}C, \\ D_M = \gamma^2I - D^TD. \end{cases} \tag{12.5.4}$$

12.5.2 H_∞范数的界与H_∞代数Riccati矩阵方程的解

命题 12.5.1 设$\gamma > 0$. 若$\|G(s)\|_\infty < \gamma$, 则$\|D\|_2 < \gamma$(或等价地$D_M > 0$).

证明 由$\|G(s)\|_\infty = \sup\limits_{\omega\in\mathbb{R}} \sigma_{\max}(G(j\omega)) = \sup\limits_{\omega\in\mathbb{R}} \lambda_{\max}(G(j\omega)^*G(j\omega))^{\frac{1}{2}}$ 得

$$\|G(s)\|_\infty \geqslant \lambda_{\max}(G(j\omega)^*G(j\omega))^{\frac{1}{2}}, \ \forall\, \omega \in \mathbb{R},$$

于是

$$\lim_{\omega\to\infty}\lambda_{\max}\left(G(j\omega)^*G(j\omega)\right)^{\frac{1}{2}}\le\|G(s)\|_\infty<\gamma.$$

这和矩阵的特征值关于其元素连续且$\lim\limits_{\omega\to\infty}(G(j\omega)^*G(j\omega))=D^TD$一起推出

$$\begin{aligned}\|D\|_2 = &\ \lambda_{\max}(D^TD)^{\frac{1}{2}}\\ = &\ \lambda_{\max}(\lim_{\omega\to\infty}G(j\omega)^*G(j\omega))^{\frac{1}{2}}\\ = &\ \lim_{\omega\to\infty}\lambda_{\max}\left(G(j\omega)^*G(j\omega)\right)^{\frac{1}{2}}<\gamma.\end{aligned}$$

显然$\|D\|_2<\gamma$的充要条件是$D_M>0$. □

命题 12.5.2 设A在虚轴上无特征根, $\gamma>0$, $\|G(s)\|_\infty<\gamma$, 则H_M在虚轴上无特征根.

证明 反证法. 若λ是H_M在虚轴上的特征根, 则λ是(A_M, iB_M)在虚轴上的不能控特征根或λ是(A_M, C_M)在虚轴上的不能观特征根. 不妨设前者成立, 则存在$0\ne y\in\mathbb{C}^n$使得

$$y^*(\lambda I-A_M)=0,\ y^*B_M=0.$$

使用命题12.5.1知$D_M>0$, 从而$y^*(\lambda I-A)=0$, 与矩阵A在虚轴上无特征根矛盾. 故H_M在虚轴上无特征根. □

推论 12.5.1 设矩阵A在虚轴上无特征根, $\gamma>0$, $D=0$, $\|G(s)\|_\infty<\gamma$, 则

$$H_{-\gamma}:=\begin{bmatrix}A & \gamma^{-2}BB^T\\ -C^TC & -A^T\end{bmatrix},$$

在虚轴上无特征根.

命题 12.5.3 设$\gamma>0$, $\|D\|_2<\gamma$(或等价地$D_M=\gamma^2I-D^TD>0$). 若P是满足矩阵方程

$$X^TA_M+A_M^TX+X^TB_MB_M^TX+C_M^TC_M=0$$

的实对称稳定解, 则$\|G(s)\|_\infty<\gamma$.

证明 记$K_\omega = (j\omega I - A)^{-1}B$, 则

$$
\begin{aligned}
& G(j\omega)^*G(j\omega) \\
= \ & (CK_\omega + D)^*(CK_\omega + D) \\
= \ & D^TD + D^TCK_\omega + K_\omega^*C^TD + K_\omega^*C^TCK_\omega \\
= \ & D^TD + D^TCK_\omega + K_\omega^*C^TD - K_\omega^*C^TDD_M^{-1}D^TCK_\omega \\
& -K_\omega^*PB_MB_M^TPK_\omega - K_\omega^*A_M^TPK_\omega - K_\omega^*PA_MK_\omega \\
= \ & \gamma^2I - (\gamma^2I - D^TD) + D^TCK_\omega + K_\omega^*C^TD \\
& -K_\omega^*C^TDD_M^{-1}D^TCK_\omega - K_\omega^*PB_MB_M^TPK_\omega \\
& +K_\omega^*(-j\omega I - A^T)PK_\omega - K_\omega^*C^TDD_M^{-1}B^TPK_\omega \\
& +K_\omega^*P(j\omega I - A)K_\omega - K_\omega^*PBD_M^{-1}D^TCK_\omega
\end{aligned}
$$

$$
\begin{aligned}
= \ & \gamma^2I - D_M + D^TCK_\omega + K_\omega^*C^TD - K_\omega^*C^TDD_M^{-1}D^TCK_\omega \\
& -K_\omega^*PB_MB_M^TPK_\omega + B^TPK_\omega - K_\omega^*C^TDD_M^{-1}B^TPK_\omega \\
& +K_\omega^*PB - K_\omega^*PBD_M^{-1}D^TCK_\omega \qquad (12.5.5) \\
= \ & \gamma^2I - \Delta^*D_M^{-1}\Delta \\
\leqslant \ & \gamma^2I, \ \forall\omega \in \mathbb{R},
\end{aligned}
$$

其中$\Delta = D_M - D^TCK_\omega - B^TPK_\omega$, 于是

$$
\sigma_{\max}(G(jw)) = \lambda_{\max}(G(jw)^*G(jw))^{\frac{1}{2}} \leqslant \gamma, \ \forall\omega \in \mathbb{R}.
$$

因而

$$
\|G(s)\|_\infty = \sup_{\omega\in\mathbb{R}} \sigma_{\max}(G(jw)) \leqslant \gamma. \tag{12.5.6}
$$

若$\|G(s)\|_\infty = \gamma$, 则$\sup\limits_{\omega\in\mathbb{R}} \sigma_{\max}(G(jw)) = \gamma$. 这和

$$
\lim_{w\to\infty} \sigma_{\max}(G(jw)) = \sigma_{\max}(D) = \|D\|_2 < \gamma
$$

一起推出存在$M > 0$使得

$$
\gamma = \sup_{\omega\in[-M,M]} \sigma_{\max}(G(jw)).
$$

使用$\sigma_{\max}(G(jw))$关于ω连续得到

$$
\gamma = \max_{\omega\in[-M,M]} \sigma_{\max}(G(jw)),
$$

并且存在$\omega_0 \in [-M, M]$使得$\gamma = \sigma_{\max}(G(jw_0))$, 即$\gamma^2$是$G(jw_0)^*G(jw_0)$的最大特征值. 设$z$是相应的特征向量, 使用式(12.5.5)得到

$$
z^*(D_M - D^TCK_{\omega_0} - B^TPK_{\omega_0})^*D_M^{-1}(D_M - D^TCK_{\omega_0} - B^TPK_{\omega_0})z = 0,
$$

于是从D_M正定推出$(D_M - D^TCK_{\omega_0} - B^TPK_{\omega_0})z = 0$, 即$z = D_M^{-1}(D^TC + B^TP)K_{\omega_0}z$, 从而$y := K_{\omega_0}z \neq 0$且$Bz = BD_M^{-1}(D^TC + B^TP)y$. 故

$$(j\omega_0 I - A)y = Bz = BD_M^{-1}(D^TC + B^TP)y,$$

于是

$$j\omega_0 y = (A_M + B_M B_M^T P)y,$$

与$A_M + B_M B_M^T P$稳定矛盾. 因而, $\|G(s)\|_\infty < \gamma$. □

定理 12.5.1 设矩阵A稳定, $\gamma > 0$, $\|G(s)\|_\infty < \gamma$, 则

(i) 矩阵方程

$$X^T A_M + A_M^T X + X^T B_M B_M^T X + C_M^T C_M = 0 \qquad (12.5.7)$$

有唯一的实对称半正定稳定解P;

(ii) 若(A, C)能观, 则$P > 0$.

证明 (i) 使用$\|G(s)\|_\infty < \gamma$和命题12.5.1得到$D_M > 0$. 由矩阵A稳定知A在虚轴上无特征根且(A, B)能稳. 使用命题12.5.2得到H_M在虚轴上无特征根. 这和(A, B)能稳、命题12.1.1和定理12.3.3一起推出矩阵方程(12.5.7)有唯一的实对称稳定解P, 于是

$$A_M^T P + PA_M + PB_M B_M^T P + C_M^T C_M = 0,$$

即

$$A^T P + P^T A + C^T C + (B^T P + D^T C)^T D_M^{-1}(B^T P + D^T C) = 0. \qquad (12.5.8)$$

注意到A稳定且

$$C^T C + (B^T P + D^T C)^T D_M^{-1}(B^T P + D^T C) \geqslant 0,$$

由定理12.1.2得$P \geqslant 0$. 故矩阵方程(12.5.7) 有唯一的实对称半正定稳定解P.

(ii) 由(i)、命题12.1.1和定理12.3.3得到. □

推论 12.5.2 设矩阵A稳定, $\gamma > 0$, $D = 0$, $\|G(s)\|_\infty < \gamma$, 则

(i) H_∞代数Riccati矩阵方程

$$A^T X + X^T A + \gamma^{-2} X^T BB^T X + C^T C = 0$$

有唯一的实对称半正定稳定解P;

(ii) 若(A, C)能观, 则$P > 0$.

12.5.3 H_∞范数的计算

下面的定理和推论给出了计算传递函数矩阵$G(s)$的H_∞范数$\|G(s)\|_\infty$的理论基础.

定理 12.5.2 设矩阵A稳定, $\gamma > 0$, $\|D\|_2 < \gamma$, 则下列说法等价:

(i) $\|G(s)\|_\infty < \gamma$;

(ii) 存在实对称正定矩阵Q使得矩阵方程

$$A_M^T X + X^T A_M + X^T B_M B_M^T X + C_M^T C_M + Q = 0$$

有唯一的实对称正定稳定解;

(iii) 存在实对称正定矩阵P使得

$$A_M^T P + P A_M + P B_M B_M^T P + C_M^T C_M < 0;$$

(iv) 存在实对称正定矩阵P使得

$$\begin{bmatrix} A_M^T P + P A_M + C_M^T C_M & P B_M \\ B_M^T P & -I \end{bmatrix} < 0.$$

证明 (i)$\Rightarrow$(ii). 由A稳定知$j\omega I - A$对任意的$\omega \in \mathbb{R}$均可逆. 使用

$$\|G(s)\|_\infty < \gamma$$

推出

$$\begin{aligned} & [(j\omega I - A)^{-1}B]^* C^T C[(j\omega I - A)^{-1}B] + [(j\omega I - A)^{-1}B]^* C^T D \\ + \ & D^T C[(j\omega I - A)^{-1}B] + D^T D < \gamma^2 I, \ \forall \omega \in \mathbb{R}, \end{aligned}$$

于是存在充分小的正数ε使得

$$\begin{aligned} & [(j\omega I - A)^{-1}B]^* (C^T C + \varepsilon I)[(j\omega I - A)^{-1}B] + [(j\omega I - A)^{-1}B]^* C^T D \\ + \ & D^T C[(j\omega I - A)^{-1}B] + D^T D < \gamma^2 I, \ \forall \omega \in \mathbb{R}, \end{aligned}$$

即

$$\begin{aligned} & [(j\omega I - A)^{-1}B]^* \hat{C}^T \hat{C}[(j\omega I - A)^{-1}B] + [(j\omega I - A)^{-1}B]^* \hat{C}^T \hat{D} \\ + \ & \hat{D}^T \hat{C}[(j\omega I - A)^{-1}B] + \hat{D}^T \hat{D} < \gamma^2 I, \ \forall \omega \in \mathbb{R}, \end{aligned}$$

其中

$$\hat{C}=\begin{bmatrix} C \\ \varepsilon^{\frac{1}{2}}I \end{bmatrix},\ \hat{D}=\begin{bmatrix} D \\ 0 \end{bmatrix}.$$

注意到A稳定且$(A,\hat{C})$能观, 由定理12.5.1知矩阵方程

$$A_M^TX+X^TA_M+X^TB_MB_M^TX+C_M^TC_M+\varepsilon I=0$$

有唯一的实对称正定稳定解. 令$Q=\varepsilon I$, 则(ii)成立.

(ii)⇒(iii). 显然.

(iii)⇒(i). 由(iii)知存在实对称正定矩阵Q使得

$$A_M^TP+PA_M+PB_MB_M^TP+C_M^TC_M+Q=0.$$

再类似于命题12.5.3的证明可推出(i)成立.

(iii)⇔(iv). 由Schur补引理得到. □

推论 12.5.3 设矩阵A稳定, $\gamma>0$, $D=0$, 则下列说法等价:

(i) $\|G(s)\|_\infty<\gamma$;

(ii) 存在实对称正定矩阵Q使得矩阵方程

$$A^TX+X^TA+\gamma^{-2}X^TBB^TX+C^TC+Q=0$$

有唯一的实对称正定稳定解;

(iii) 存在实对称正定矩阵P使得

$$A^TP+PA+\gamma^{-2}PBB^TP+C^TC<0;$$

(iv) 存在实对称正定矩阵P使得

$$\begin{bmatrix} A^TP+PA+C^TC & PB \\ B^TP & -\gamma^2I \end{bmatrix}<0.$$

根据定理12.5.2, 可给出计算$\|G(s)\|_\infty$的迭代方法如下:

算法 12.5.1 (计算$\|G(s)\|_\infty$的近似值) 对于$a=\sigma_{\max}(G(i\infty))$, 充分小的正整数$\varepsilon$和满足$\|G(s)\|_\infty<b$的充分大的正数$b$, 执行下面的步骤:

(i) $\gamma=\frac{a+b}{2}$, 并根据定理12.5.2(或推论12.5.3)判断是否$\|G(s)\|_\infty<\gamma$.

(ii) 若$\|G(s)\|_\infty < \gamma$, 则令$b = \gamma$; 否则, 令$a = \gamma$.

(iii) 若$|a - b| > \varepsilon$, 则返回(i); 否则算法结束.

由上面的算法得到的$\|G(s)\|_\infty$满足$\|G(s)\|_\infty \in [\gamma - \varepsilon, \gamma + \varepsilon]$. 应该强调的是(i)可通过MATLAB的LMI工具箱来实现.

习 题

1. 证明: λ是(A, B)的不能控特征根的充要条件是存在非零的$v \in \mathbb{C}^n$使得

$$\begin{bmatrix} A & -\eta BB^T \\ -M & -A^T \end{bmatrix} \begin{bmatrix} 0 \\ v \end{bmatrix} = -\lambda \begin{bmatrix} 0 \\ v \end{bmatrix}, \ \eta = \pm 1.$$

2. 证明: λ是(A, C)的不能观特征根的充要条件是存在非零的$v \in \mathbb{C}^n$使得

$$\begin{bmatrix} A & -\eta BB^T \\ -M & -A^T \end{bmatrix} \begin{bmatrix} v \\ 0 \end{bmatrix} = \lambda \begin{bmatrix} v \\ 0 \end{bmatrix}, \ \eta = \pm 1.$$

3. 设A是n阶矩阵, 证明: A的特征值均具有负实部的充要条件是存在$X > 0$使得$AX + XA^* < 0$.

   (提示: 证必要性时考虑Lyapunov方程$AX + XA^* = -I$.)

4. 设B和C是给定的具有适当维数的实矩阵, γ是给定的正数, 则$\|CB\|_2 < \gamma$的充要条件是H_∞代数Riccati矩阵方程$2Y + \gamma^{-2}YBB^TY + C^TC = 0$有唯一的实对称解.

5. 设B和C是任意给定的具有适当维数的实矩阵, $\gamma > 0$是给定的常数. 若实对称矩阵Y满足方程$2Y + \gamma^{-2}YBB^TY + C^TC = 0$, 则$Y$是半负定的.

6. 证明: 矩阵方程$-2X - X^TBB^TX + C^TC = 0$有唯一的实对称半正定稳定解.

7. 证明: 矩阵方程$2X - X^TBB^TX + C^TC = 0$有唯一的实对称半负定解.

8. 设$\gamma > 0$, A在虚轴上无特征根, (A, B)能稳, 则

$$\|C(sI - A)^{-1}B + D\|_\infty < \gamma$$

   的充要条件是下面(i)和(ii)成立:

   (i) $\|D\|_2 < \gamma$ (或$D_M = \gamma^2 I - D^TD > 0$);

   (ii) H_∞代数Riccati矩阵方程

$$X^TA_M + A_M^TX + X^TB_MB_M^TX + C_M^TC_M = 0$$

   有唯一的实对称稳定解, 其中

$$A_M = A + BD_M^{-1}D^TC, \ B_M = BD_M^{-\frac{1}{2}},$$

$$C_M = (I + DD_M^{-1}D^T)^{\frac{1}{2}}C, \ D_M = \gamma^2 I - D^TD.$$

9. 设矩阵A在虚轴上无特征根, $\gamma > 0$, $\|C(sI - A)^{-1}B + D\|_\infty < \gamma$, 证明: $\|T(s)\|_\infty < 1$, 其中$T(s) = C_M(sI - A_M)^{-1}B_M$, A_M, B_M和C_M同上题中定义.

10. 设(A,B)能稳, (A,C)能检测, γ是给定的正数, 证明: 若

$$\|C(sI-A)^{-1}B+D\|_\infty<\gamma,$$

则H_2代数Riccati矩阵方程

$$PA_M+A_M^TP-PB_MB_M^TP+C_M^TC_M=0$$

有唯一的实对称半正定稳定解, 其中$T(s)=C_M(sI-A_M)^{-1}B_M$, A_M, B_M和C_M同上题中定义.

11. 设A在虚轴上无特征根, (A,B)能稳, $\gamma>0$, 则下列说法等价:

(i) $\|C(sI-A)^{-1}B\|_\infty<\gamma$;

(ii) H_∞代数Riccati矩阵方程

$$A^TX+X^TA+\gamma^{-2}X^TBB^TX+C^TC=0$$

有唯一的实对称稳定解.

参 考 文 献

[1] 张显. 关于秩的两个等式的充要条件的矩阵证法. *纺织高校基础科学学报*, 12(1):75–76, 1999.

[2] 曹重光. *线性代数*. 内蒙古科学技术出版社, 内蒙古, 1999.

[3] 王萼芳, 石生明. *高等代数(第三版)*. 高等教育出版社, 北京, 2003.

[4] J. M. Ortega. *Matrix Theory: A Second Course*. Plenum Press, New York, 1987.

[5] P. Lancaster. *Theory of Matrices*. Academic Press, New York, 1969.

[6] G. R. Duan. *Analysis and Design of Descriptor Linear Systems*, volume 23 of *Advances in Mechanics and Mathematics*. Springer, New, York Dordrecht, Heidelberg, London, 2010.

[7] 郑大钟. *线性系统理论(第二版)*. 清华大学出版社, 北京, 2002.

[8] X. Zhang, X. M. Tang, and C. G. Cao. *Preserver Problems on Spaces of Matrices*. Science Press, Beijing, 2007.

[9] F. R. Gantmacher. *Applications of the Theory of Matrices*. Interscience, New York, 1959.

[10] G.-R. Duan and X. Zhang. Dynamical order assignment in linear descriptor systems via state derivative feedback. In *Proc. 41st IEEE Conf. Decision Control*, pages 4533–4538, 2002.

[11] X. Zhang. The general common Hermitian nonnegative-definite solution to the matrix equations $AXA^* = B$ and $CXC^* = D$. *Linear & Multilinear Algebra*, 52(1):49–60, 2004.

[12] 程其襄. *实变函数与泛函分析基础*. 高等教育出版社, 北京, 1983.

[13] 刘炳初. *泛函分析(第二版)*. 科学出版社, 北京, 2004.

[14] A. R. Amir-Moéz. Extreme properties of eigenvalues of a Hermitian transformation and singular values of the sum and product of linear transformations. *Duke Math. J.*, 23(3):463–476, 1956.

[15] C. G. Cao, X. Zhang, and X. M. Tang. Reverse order law of group inverses of products of two matrices. *Appl. Math. Comput.*, 158:489–495, 2004.

[16] T. Iwasaki and R. E. Skelton. All controllers for the general H_∞ control problem: LMI existence conditions and state space formulas. *Automatica*, 30(8):1307–1317, 1994.

[17] T. Iwasaki and G. Shibata. LPV system analysis via quadratic separator for uncertain implicit systems. *IEEE Trans. Autom. Control*, 46(8):1195–1208, 2001.

[18] P. Gahinet, P. Apkarian, and M. Chilali. Affine parameter-dependent Lyapunov functions and real parametric uncertainty. *IEEE Trans. Autom. Control*, 41(3):436–442, 1996.

[19] D. Yue, E. G. Tian, and Y. J. Zhang. A piecewise analysis method to stability analysis of linear continuous/discrete systems with time-varying delay. *Int. J. Robust Nonlinear Control*, 19(13):1493–1518, 2009.

[20] L. H. Xie. Output feedback H_∞ control of systems with parameter uncertainty. *Int. J. Control*, 63(4):741–750, 1996.

[21] K. Gu. An integral inequality in the stability problem of time-delay systems. In *Proc. 39th IEEE Conf. Decision Control*, pages 2805–2810, 2000.

[22] X. Jiang, Q. L. Han, and X. Yu. Stability criteria for linear discrete-time systems with interval-like time-varying delay. In *Proc. 2005 Amer. Control Conf.*, pages 2817–2822, 2005.

[23] Q.-L. Han and D. Yue. Absolute stability of Lur'e systems with time-varying delay. *IET Control Theory Appl.*, 1(3):854–859, 2007.

[24] H. J. Wang, A. K. Xue, R. Q. Lu, and J. H. Wang. Delay-dependent robust stability and stabilization for uncertain singular system with time-varying delay. In *Proc. 2008 Amer. Control Conf.*, pages 3626–3631, 2008.

附录 A　定理3.1.1的证明

定理3.1.1的证明可以通过组合下面的几个命题得到.

命题 A.0.1　设M, $N\in\mathbb{R}^{m\times n}$, 则存在$P\in GL_m(\mathbb{R})$和$Q\in GL_n(\mathbb{R})$使得

$$P(sM-N)Q=\mathrm{diag}(0_{m_0\times n_0}, sM_0-N_0),\ \forall s\in\mathbb{C}, \tag{A.0.1}$$

并且$\begin{bmatrix} M_0 & N_0 \end{bmatrix}$行满秩、$\begin{bmatrix} M_0 \\ N_0 \end{bmatrix}$列满秩.

证明　设Q_1, $\cdots$, Q_{n_0}是$\mathcal{N}\left(\begin{bmatrix} M \\ N \end{bmatrix}\right)$的基, 将其扩充为$\mathbb{R}^n$的基$Q_1$, $\cdots$, $Q_{n_0}, Q_{n_0+1}, \cdots, Q_n$; 设$P_1$, $\cdots$, P_{m_0}是$\mathcal{N}\left(\begin{bmatrix} M^T \\ N^T \end{bmatrix}\right)$的基, 将其扩充为$\mathbb{R}^m$的基$P_1, \cdots, P_{m_0}, P_{m_0+1}, \cdots, P_m$. 令

$$Q=\begin{bmatrix} Q_1 & Q_2 & \cdots & Q_n \end{bmatrix}, P=\begin{bmatrix} P_1 & P_2 & \cdots & P_m \end{bmatrix}^T,$$

则容易验证式(A.0.1)成立.

设z是$\begin{bmatrix} M_0 \\ N_0 \end{bmatrix}x=0$的任意解, 则

$$M_0z=0,\ N_0z=0.$$

利用式(A.0.1)推出$Q\begin{bmatrix} 0 \\ z \end{bmatrix}$是

$$\begin{bmatrix} M \\ N \end{bmatrix}x=0$$

的解, 于是$\begin{bmatrix} Q_{n_0+1} & \cdots & Q_n \end{bmatrix}z$可由$Q_1$, $\cdots$, Q_{n_0}线性表出, 这与Q_1, $\cdots$, Q_{n_0}, Q_{n_0+1}, $\cdots$, Q_n是$\mathbb{R}^n$的基一起推出$z=0$, 故$\begin{bmatrix} M_0 \\ N_0 \end{bmatrix}$列满秩. 同理, $\begin{bmatrix} M_0 & N_0 \end{bmatrix}$行满秩. □

命题 A.0.2 若$M,\ N\in\mathbb{R}^{m\times n}$满足$\begin{bmatrix} M \\ N \end{bmatrix}$列满秩并且

$$\operatorname{rank}(sM-N)<n,\ \forall s\in\mathbb{C},\tag{A.0.2}$$

则:

(i) 存在正整数k和

$$x_k(s):=x_0+x_1s+\cdots+x_ks^k,\ x_i\in\mathbb{R}^n, i=0,1,2,\cdots,k, x_k\neq 0,\tag{A.0.3}$$

使得

$$(sM-N)x_k(s)=0,\ \forall s\in\mathbb{C};\tag{A.0.4}$$

(ii) ε是使得式(A.0.4)成立的最小正整数k的充要条件是

$$\Phi_\varepsilon\text{列降秩},\ \Phi_i\text{列满秩},\quad i=0,1,2,\cdots,\varepsilon-1,\tag{A.0.5}$$

其中

$$\Phi_k=\begin{bmatrix} -N & & & & \\ M & -N & & & \\ & M & -N & & \\ & & \ddots & \ddots & \\ & & & M & -N \\ & & & & M \end{bmatrix}\in\mathbb{R}^{(k+2)m\times(k+1)n},$$

$$k=0,1,2,\cdots,\varepsilon;$$

(iii) 当ε是使得式(A.0.4)成立的最小正整数时, $Nx_1, Nx_2,\cdots,Nx_\varepsilon$线性无关, 并且$x_0,x_1,\cdots,x_\varepsilon$线性无关;

(iv) 当ε是满足式(A.0.4)的最小正整数时, 存在可逆阵P和Q使得

$$P(sM-N)Q=\operatorname{diag}(L_\varepsilon(s), sM_1-N_1),\ \forall s\in\mathbb{C}.\tag{A.0.6}$$

证明 (i) 存在可逆的多项式矩阵$P(s)$和$Q(s)$使得

$$P(s)(sM-N)Q(s)=\operatorname{diag}(d_1(s),\cdots,d_r(s),0,\cdots,0),$$

其中$d_1(s), \cdots, d_r(s)$是非零的实系数多项式, $r = \max\limits_{s\in\mathbb{C}} \operatorname{rank}(sM - N)$. 注意到$Q^{-1}(s)$仍是多项式矩阵, $Q^{-1}(s)$的最后一列可表示成形式

$$y_l(s) = y_0 + y_1 s + \cdots + y_l s^l,\ y_i \in \mathbb{R}^n,\quad i = 0, 1, 2, \cdots, l,\ y_l \neq 0.$$

使用式(A.0.2)推出$(sM - N)sy_l(s) = 0,\ \forall s \in \mathbb{C}$.

(ii) 设ε是正整数, 易见

$$\begin{aligned}
& (sM - N)x_\varepsilon(s) = 0,\ \forall s \in \mathbb{C} \\
\Leftrightarrow\ & -Nx_0 + s(Mx_0 - Nx_1) + \cdots \\
& + s^\varepsilon(Mx_{\varepsilon-1} - Nx_\varepsilon) + s^{\varepsilon+1}Mx_\varepsilon = 0,\ \forall s \in \mathbb{C} \\
\Leftrightarrow\ & \begin{cases} -Nx_0 = 0, \\ Mx_0 - Nx_1 = 0, \\ Mx_1 - Nx_2 = 0, \\ \qquad \vdots \\ Mx_{\varepsilon-1} - Nx_\varepsilon = 0, \\ Mx_\varepsilon = 0. \end{cases} \\
\Leftrightarrow\ & \Phi_\varepsilon \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_{\varepsilon-1} & x_\varepsilon \end{bmatrix}^T = 0.
\end{aligned} \tag{A.0.7}$$

由$\begin{bmatrix} M \\ N \end{bmatrix}$列满秩知(ii)成立.

(iii) 若$Nx_1 = 0$, 则由式(A.0.7)得$\Phi_{\varepsilon-1}(x_1 + x_2 s + \cdots + x_\varepsilon s^{\varepsilon-1}) = 0$. 使用(ii)及$\varepsilon$的最小性可推出$x_1 = \cdots = x_\varepsilon = 0$, 与$x_\varepsilon \neq 0$矛盾. 故$Nx_1 \neq 0$. 因而, 若$Nx_1, Nx_2, \cdots, Nx_\varepsilon$线性相关, 则存在$h$ $(1 < h \leqslant \varepsilon)$及$\alpha_1, \cdots, \alpha_{h-1} \in \mathbb{R}$使得

$$Nx_h = \alpha_1 Nx_1 + \alpha_2 Nx_2 + \cdots + \alpha_{h-1} Nx_{h-1}.$$

由式(A.0.7)知$Mx_{h-1} = \alpha_1 Mx_0 + \alpha_2 Mx_1 + \cdots + \alpha_{h-1} Mx_{h-2}$, 即$M(x_{h-1} - \alpha_1 x_0 - \cdots - \alpha_{h-1} x_{h-2}) = 0$. 记$x_{h-1}^* = x_{h-1} - \alpha_1 x_0 - \cdots - \alpha_{h-1} x_{h-2}$, 则

$$Mx_{h-1}^* = 0, \tag{A.0.8}$$

并且

$$\begin{aligned} Nx_{h-1}^* = &\ N(x_{h-1} - \alpha_1 x_0 - \cdots - \alpha_{h-1} x_{h-2}) \\ = &\ M(x_{h-2} - \alpha_2 x_0 - \cdots - \alpha_{h-1} x_{h-3}). \end{aligned}$$

记$x_{h-2}^* = x_{h-2} - \alpha_2 x_0 - \cdots - \alpha_{h-1} x_{h-3}$, 则

$$Mx_{h-2}^* = Nx_{h-1}^*. \tag{A.0.9}$$

同理, 存在$x_{h-3}^*, \cdots, x_1^*, x_0^*$使得

$$Nx_0^* = 0,\ Mx_i^* = Nx_{i+1}^*,\ i = 0, 1, \cdots, h-3. \tag{A.0.10}$$

令$x^*(s) = x_0^* + x_1^* s + \cdots + x_{h-1}^* s^{h-1}$, 则由式(A.0.8)至式(A.0.10)推出

$$(sM - N)x^*(s) = 0,$$

但$h-1<\varepsilon$. 这与ε是使得式(A.0.4)成立的最小正整数矛盾. 所以Nx_1, Nx_2, $\cdots$, Nx_ε线性无关.

现在往证$x_0,x_1,\cdots,x_\varepsilon$线性无关. 令$\alpha_0,\alpha_1,\cdots,\alpha_\varepsilon\in\mathbb{R}$满足

$$\alpha_0x_0+\alpha_1x_1+\cdots+\alpha_\varepsilon x_\varepsilon=0,$$

则

$$\alpha_0Nx_0+\alpha_1Nx_1+\cdots+\alpha_\varepsilon Nx_\varepsilon=0.$$

由$Nx_0=0$及$Nx_1,Nx_2,\cdots,Nx_\varepsilon$线性无关知$\alpha_1=\cdots=\alpha_\varepsilon=0$, 从而$\alpha_0x_0=0$. 若$x_0=0$, 则$Nx_1=Mx_0=0$, 于是$Nx_1,Nx_2,\cdots,Nx_\varepsilon$线性相关, 矛盾. 故$x_0\neq 0$, 所以$\alpha_0=0$. 总之, $x_0,x_1,\cdots,x_\varepsilon$线性无关.

(iv) 将$Nx_1,Nx_2,\cdots,Nx_\varepsilon$扩充为$\mathbb{R}^m$的基

$$Nx_1,Nx_2,\cdots,Nx_\varepsilon,y_1,\cdots,y_{m-\varepsilon};$$

将$x_0,x_1,\cdots,x_\varepsilon$扩充为$\mathbb{R}^n$的基$x_0,x_1,\cdots,x_\varepsilon,z_1,\cdots,z_{n-(\varepsilon+1)}$. 令

$$P_1=\begin{bmatrix}Nx_1 & \cdots & Nx_\varepsilon & y_1 & \cdots & y_{m-\varepsilon}\end{bmatrix}^{-1},$$

$$Q_1=\begin{bmatrix}x_0 & x_1 & \cdots & x_\varepsilon & z_1 & \cdots & z_{n-(\varepsilon+1)}\end{bmatrix},$$

则

$$\begin{aligned}
&P_1(sM-N)Q_1\\
=\ &P_1\begin{bmatrix}(sM-N)x_0 & (sM-N)x_1 & \cdots & (sM-N)x_\varepsilon & \star & \cdots & \star\end{bmatrix}\\
=\ &P_1\begin{bmatrix}sNx_1 & sNx_2-Nx_1 & \cdots & sNx_\varepsilon-Nx_{\varepsilon-1} & -Nx_\varepsilon & \star & \cdots & \star\end{bmatrix}\\
=\ &\begin{bmatrix}Nx_1 & \cdots & Nx_\varepsilon & y_1 & \cdots & y_{m-\varepsilon}\end{bmatrix}^{-1}\\
&\times\begin{bmatrix}sNx_1 & sNx_2-Nx_1 & \cdots & sNx_\varepsilon-Nx_{\varepsilon-1} & -Nx_\varepsilon & \star & \cdots & \star\end{bmatrix}\\
=\ &\begin{bmatrix}L_\varepsilon(s) & sM_{01}-N_{01}\\ 0 & sM_1-N_1\end{bmatrix},\ \forall s\in\mathbb{C},
\end{aligned}\tag{A.0.11}$$

这里$\star$表示后面不被使用的块(下同).

现在只需证明存在实矩阵X和Y满足

$$\begin{bmatrix}I & X\\ 0 & I\end{bmatrix}\begin{bmatrix}L_\varepsilon(s) & sM_{01}-N_{01}\\ 0 & sM_1-N_1\end{bmatrix}\begin{bmatrix}I & Y\\ 0 & I\end{bmatrix}=\mathrm{diag}(L_\varepsilon(s),sM_1-N_1).\tag{A.0.12}$$

通过直接计算易见: 式(A.0.12)成立的充要条件是

$$L_\varepsilon(s)Y+sM_{01}-N_{01}+X(sM_1-N_1)=0.\tag{A.0.13}$$

因而, 只需证存在实矩阵X和Y满足式(A.0.13).

将X, Y, M_{01}和N_{01}按行分块有

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_\varepsilon \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{\varepsilon+1} \end{bmatrix}, \quad M_{01} = \begin{bmatrix} M_{01}^{(1)} \\ M_{01}^{(2)} \\ \vdots \\ M_{01}^{(\varepsilon)} \end{bmatrix}, \quad N_{01} = \begin{bmatrix} N_{01}^{(1)} \\ N_{01}^{(2)} \\ \vdots \\ N_{01}^{(\varepsilon)} \end{bmatrix},$$

于是式(A.0.13)等价于

$$sY_j - Y_{j+1} + sM_{01}^{(j)} - N_{01}^{(j)} + sX_jM_1 - X_jN_1 = 0, \ j = 1, 2, \cdots, \varepsilon.$$

由s的任意性得

$$\begin{cases} Y_j + M_{01}^{(j)} + X_jM_1 = 0, \\ Y_{j+1} + N_{01}^{(j)} + X_jN_1 = 0, \end{cases} \quad j = 1, 2, \cdots, \varepsilon,$$

即

$$\begin{cases} Y_1 = -M_{01}^{(1)} - X_1M_1, \\ Y_k = -M_{01}^{(k)} - X_kM_1 = -N_{01}^{(k-1)} - X_{k-1}N_1, \ k = 2, \cdots, \varepsilon, \\ Y_{\varepsilon+1} = -N_{01}^{(\varepsilon)} - X_\varepsilon N_1. \end{cases}$$

因而, 只需证存在实矩阵X满足

$$M_{01}^{(k)} + X_kM_1 = N_{01}^{(k-1)} + X_{k-1}N_1, \ k = 2, \cdots, \varepsilon. \qquad \text{(A.0.14)}$$

既然式(A.0.14)等价于

$$\begin{bmatrix} X_1 & X_2 & \cdots & X_\varepsilon \end{bmatrix} \Psi = \begin{bmatrix} N_{01}^{(1)} - M_{01}^{(2)} & \cdots & N_{01}^{(\varepsilon-1)} - M_{01}^{(\varepsilon)} \end{bmatrix},$$

其中

$$\Psi = \begin{bmatrix} -N_1 & & & \\ M_1 & -N_1 & & \\ & M_1 & \ddots & \\ & & \ddots & -N_1 \\ & & & M_1 \end{bmatrix},$$

故只需证Ψ列满秩.

由式(A.0.11)知存在置换矩阵P_2和Q_2满足

$$P_2\Phi_{\varepsilon-1}Q_2 = \begin{bmatrix} C & \star \\ 0 & \Psi \end{bmatrix},$$

其中$C \in \mathbb{R}^{(\varepsilon+1)\varepsilon\times(\varepsilon+1)\varepsilon}$. 注意到$\varepsilon$是使得式(A.0.4)成立的最小正整数, 我们从(ii)得到$\Phi_{\varepsilon-1}$列满秩. 所以C可逆且Ψ列满秩. □

命题 A.0.3 若$M, N \in \mathbb{R}^{m\times n}$满足$\begin{bmatrix} M & N \end{bmatrix}$行满秩并且

$$\operatorname{rank}(sM - N) < m, \ \forall s \in \mathbb{C},$$

则存在正整数k和$P \in GL_m(\mathbb{R})$、$Q \in GL_n(\mathbb{R})$使得

$$P(sM - N)Q = \operatorname{diag}((L_k(s))^T, sM_1 - N_1), \ \forall s \in \mathbb{C}.$$

证明 类似于命题A.0.2. □

命题 A.0.4 设$M, N \in \mathbb{R}^{n\times n}$, 若存在$s \in \mathbb{C}$使得$\det(sM - N) \neq 0$, 则存在$P, Q \in GL_n(\mathbb{R})$使得

$$P(sM - N)Q = \operatorname{diag}(sI_{n_1} - N_1, sM_2 - I_{n_2}), \ \forall s \in \mathbb{C},$$

其中$N_1 \in \mathbb{R}^{n_1\times n_1}$, $M_2 \in \mathbb{R}^{n_2\times n_2}$, 且$M_2$为幂零阵.

证明 易见, 存在实数s_0使得$\det(s_0 M - N) \neq 0$. 令

$$\overline{M} = (s_0 M - N)^{-1} M, \ \overline{N} = (s_0 M - N)^{-1} N,$$

则

$$s_0 \overline{M} - \overline{N} = I. \tag{A.0.15}$$

由$\overline{M}$的Fitting分解知存在可逆阵$Q \in \mathbb{R}^{n\times n}$使得

$$Q^{-1}\overline{M}Q = \begin{bmatrix} \overline{M}_1 & 0 \\ 0 & \overline{M}_2 \end{bmatrix},$$

其中$\overline{M}_1 \in \mathbb{R}^{n_1\times n_1}$为可逆阵, $\overline{M}_2 \in \mathbb{R}^{n_2\times n_2}$为幂零矩阵. 由式(A.0.15)知

$$Q^{-1}\overline{N}Q = Q^{-1}(s_0\overline{M} - I)Q = \begin{bmatrix} s_0\overline{M}_1 - I & 0 \\ 0 & s_0\overline{M}_2 - I \end{bmatrix} := \begin{bmatrix} \overline{N}_1 & 0 \\ 0 & \overline{N}_2 \end{bmatrix}. \tag{A.0.16}$$

显然, $\overline{N}_1 \in \mathbb{R}^{n_1\times n_1}$, $\overline{N}_2 \in \mathbb{R}^{n_2\times n_2}$可逆. 从而有

$$\begin{bmatrix} \overline{M}_1^{-1} & 0 \\ 0 & \overline{N}_2^{-1} \end{bmatrix} Q^{-1}\overline{M}Q = \begin{bmatrix} I & 0 \\ 0 & \overline{N}_2^{-1}\overline{M}_2 \end{bmatrix}$$

且

$$\begin{bmatrix} \overline{M}_1^{-1} & 0 \\ 0 & \overline{N}_2^{-1} \end{bmatrix} Q^{-1}\overline{N}Q = \begin{bmatrix} \overline{M}_1^{-1}\overline{N}_1 & 0 \\ 0 & I \end{bmatrix}.$$

令$M_2=\overline{N}_2^{-1}\overline{M}_2$, $N_1=\overline{M}_1^{-1}\overline{N}_1$及

$$P=\begin{bmatrix}\overline{M}_1^{-1} & 0\\ 0 & \overline{N}_2^{-1}\end{bmatrix}Q^{-1}(s_0M-N)^{-1},$$

则只需证M_2是幂零矩阵. 事实上, 由$\overline{M}_2$为幂零矩阵知存在正整数h满足$\overline{M}_2^h=0$, 这和$M_2=\overline{N}_2^{-1}\overline{M}_2=(s_0\overline{M}_2-I)^{-1}\overline{M}_2=\overline{M}_2(s_0\overline{M}_2-I)^{-1}$一起推出$M_2^h=\overline{N}_2^{-h}\overline{M}_2^h=0$, 即$M_2$为幂零矩阵. □

附录 B 定理3.3.1的证明

记$B = \begin{bmatrix} b_1 & b_2 & \cdots & b_m \end{bmatrix}$, 其中$b_i \in \mathbb{R}^n$, $i = 1, 2, \cdots, m$. 由(A, B)能控知$\text{rank} C_{n,A,B} = n$. 这与$B$列满秩一起推出: 可在矩阵$C_{n,A,B}$的列中选择出$n$个线性无关的向量$x_1 (= b_1)$, $\cdots$, $x_m (= b_m)$, x_{m+1}, $\cdots$, x_n, 并且使得在矩阵$C_{n,A,B}$的列中排在x_s左边的向量一定可由向量x_1, x_2, $\cdots$, x_{s-1}线性表出, $s = m+1, m+2, \cdots, n$. 由矩阵$C_{n,A,B}$的结构知, 这n个向量可表示成形式

$$b_1, Ab_1, \cdots, A^{n_1 - 1} b_1; b_2, Ab_2, \cdots, A^{n_2 - 1} b_2; \cdots; b_m, Ab_m, \cdots, A^{n_m - 1} b_m,$$

其中n_1, n_2, $\cdots$, n_m均为正整数且满足$n_1 + n_2 + \cdots + n_m = n$, 于是矩阵

$$Q := \begin{bmatrix} C_{n_1,A,b_1} & C_{n_2,A,b_2} & \cdots & C_{n_m,A,b_m} \end{bmatrix}$$

可逆. 令

$$Q^{-1} = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_m \end{bmatrix}, \ Q_i = \begin{bmatrix} * & \cdots & * & q_i \end{bmatrix}^T, \ i = 1, 2, \cdots, m,$$

其中$Q_i \in \mathbb{R}^{n_i \times n}$, $q_i \in \mathbb{R}^n$. 由$Q^{-1} Q = I_n$得下面的两个结论:

(i) 若$A^k b_j$是矩阵Q的列, 则$k < n_j$且

$$q_i^T A^k b_j = \begin{cases} 1, & \text{若}(k, j) = (n_i - 1, i), \\ 0, & \text{否则}, \end{cases} \quad i, j = 1, 2, \cdots, m;$$

(ii) 对于$i = 1, 2, \cdots, m$, 若z是矩阵$C_{n,A,B}$中排在$A^{n_i - 1} b_i$前面的列的线性组合, 则$q_i^T z = 0$.

设

$$G_i = \begin{bmatrix} A^{n_i - 1} b_i & \cdots & Ab_i & b_i \end{bmatrix}, W_i = \begin{bmatrix} q_i^T \\ q_i^T A \\ \vdots \\ q_i^T A^{n_i - 1} \end{bmatrix}, \ i = 1, 2, \cdots, m,$$

$$W = \begin{bmatrix} W_1^T & W_2^T & \cdots & W_m^T \end{bmatrix}^T,$$

则由结论(i)得

$$W_iG_i = \begin{bmatrix} 1 & & & \\ & \ddots & 0 & \\ & * & \ddots & \\ & & & 1 \end{bmatrix}, \ i = 1, 2, \cdots, m,$$

$$W_iG_j = \begin{bmatrix} 0 & & & \\ & \ddots & 0 & \\ & * & \ddots & \\ & & & 0 \end{bmatrix}, \ i, j = 1, 2, \cdots, m, i \neq j.$$

使用

$$W \begin{bmatrix} G_1 & G_2 & \cdots & G_m \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_m \end{bmatrix} \begin{bmatrix} G_1 & G_2 & \cdots & G_m \end{bmatrix}$$

推出W可逆. 记$P = W^{-1}$, 则由结论(i)和(ii)得证.

附录 C　定理3.4.4的证明

由奇异值分解, 存在正交阵U_1, V_1和正定对角阵$D_1 \in \mathbb{R}^{n_0\times n_0}$使得

$$E = U_1 \mathrm{diag}(D_1, 0) V_1^T. \tag{C.0.1}$$

令

$$U_1^T B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \tag{C.0.2}$$

其中$B_1 \in \mathbb{R}^{n_0\times r}$, $B_2 \in \mathbb{R}^{(n-n_0)\times r}$. 组合式(C.0.1)和式(C.0.2)推出

$$\begin{aligned}
& \mathrm{rank}\begin{bmatrix} E & B \end{bmatrix} \\
= & \mathrm{rank}\begin{bmatrix} U_1 \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} V_1^T & U_1 \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \end{bmatrix} \\
= & \mathrm{rank}\begin{bmatrix} D_1 & 0 & B_1 \\ 0 & 0 & B_2 \end{bmatrix} \\
= & n_0 + \mathrm{rank} B_2.
\end{aligned} \tag{C.0.3}$$

若$B_2 = 0$, 则由式(C.0.3)得$n_0 = \mathrm{rank}\begin{bmatrix} E & B \end{bmatrix} = n_1 + r$. 由奇异值分解, 存在正交阵$\widehat{U}$, $\widehat{V}$和正定对角阵$\widehat{D} \in \mathbb{R}^{r\times r}$使得

$$D_1^{-1} B_1 = \widehat{U} \begin{bmatrix} 0 \\ \widehat{D} \end{bmatrix} \widehat{V}^T.$$

令

$$T = \widehat{V}^T,\ V = \mathrm{diag}(\widehat{U}^T, I_{n-n_0}) V_1^T,$$

$$N = U_1 \mathrm{diag}(D_1 \widehat{U}, I_{n-n_0}) \mathrm{diag}(I_{n_1}, \widehat{D}, I_{n-n_0}),$$

$$M = \begin{bmatrix} \widehat{D}^{-1} & 0_{r\times(n-n_1-r)} \end{bmatrix},$$

则

$$
\begin{aligned}
& N\begin{bmatrix} I_{n_1} & 0 \\ 0 & M \\ 0 & 0 \end{bmatrix} V \\
= \ & U_1\mathrm{diag}(D_1\widehat{U}, I)\mathrm{diag}(I_{n_1}, \widehat{D}, I_{n-n_0})\mathrm{diag}(I_{n_1}, \widehat{D}^{-1}, 0)\mathrm{diag}(\widehat{U}^T, I)V_1^T \\
= \ & U_1\mathrm{diag}(D_1\widehat{U}, I)\mathrm{diag}(I_{n_0}, 0)\mathrm{diag}(\widehat{U}^T, I)V_1^T \\
= \ & U_1\mathrm{diag}(D_1, 0)V_1^T \\
= \ & E
\end{aligned}
$$

且

$$
\begin{aligned}
N\begin{bmatrix} 0_{n_1\times r} \\ I_r \\ 0 \end{bmatrix} T = \ & U_1\mathrm{diag}(D_1\widehat{U}, I)\mathrm{diag}(I_{n_1}, \widehat{D}, I_{n-n_0})\begin{bmatrix} 0_{n_1\times r} \\ I_r \\ 0 \end{bmatrix}\widehat{V}^T \\
= \ & U_1\mathrm{diag}(D_1\widehat{U}, I)\begin{bmatrix} 0_{n_1\times r} \\ \widehat{D} \\ 0 \end{bmatrix}\widehat{V}^T \\
= \ & U_1\begin{bmatrix} D_1\widehat{U}\begin{bmatrix} 0_{n_1\times r} \\ \widehat{D} \end{bmatrix}\widehat{V}^T \\ 0 \end{bmatrix} \\
= \ & U_1\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \\
= \ & B,
\end{aligned}
$$

即式(3.4.3)成立.

若$B_2 \neq 0$且$0 < \mathrm{rank}B_2 < r$, 则$n_0 = n_1$. 由奇异值分解知存在正交阵U_2, V_2和正定对角阵$D_2 \in \mathbb{R}^{r\times r}$满足

$$B_2 = U_2\begin{bmatrix} D_2 \\ 0 \end{bmatrix} V_2^T.$$

令

$$T = V_2^T,\ V = V_1^T,\ M = 0,$$

$$N = U_1\mathrm{diag}(I_{n_0}, U_2)\left[\begin{array}{c|cc} D_1 & B_1V_2 & 0 \\ \hline 0 & D_2 & 0 \\ 0 & 0 & I \end{array}\right],$$

则

$$
\begin{aligned}
& N\begin{bmatrix} I_{n_1} & 0 \\ 0 & M \\ 0 & 0 \end{bmatrix} V \\
= \ & U_1 \text{diag}(I_{n_0}, U_2) \left[\begin{array}{c|cc} D_1 & B_1V_2 & 0 \\ \hline 0 & D_2 & 0 \\ 0 & 0 & I \end{array}\right] \text{diag}(I_{n_0}, 0)V_1^T \\
= \ & U_1 \text{diag}(D_1, 0)V_1^T \\
= \ & E
\end{aligned}
$$

且

$$
\begin{aligned}
N\begin{bmatrix} 0_{n_1\times r} \\ I_r \\ 0 \end{bmatrix} T = \ & U_1 \text{diag}(I_{n_0}, U_2) \left[\begin{array}{c|cc} D_1 & B_1V_2 & 0 \\ \hline 0 & D_2 & 0 \\ 0 & 0 & I \end{array}\right] \left[\begin{array}{c} 0_{n_1\times r} \\ \hline I_r \\ 0 \end{array}\right] V_2^T \\
= \ & U_1 \text{diag}(I_{n_0}, U_2) \left[\begin{array}{c} B_1V_2 \\ \hline D_2 \\ 0 \end{array}\right] V_2^T \\
= \ & U_1 \begin{bmatrix} B_1 \\ U_2 \begin{bmatrix} D_2 \\ 0 \end{bmatrix} V_2^T \end{bmatrix} \\
= \ & U_1 \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \\
= \ & B,
\end{aligned}
$$

即式(3.4.3)成立.

若$B_2 \neq 0$且$0 < \text{rank}B_2 < r$, 则由式(C.0.3)得$n_0 \neq n_1$. 由奇异值分解, 存在正交阵U_2, V_2和$r + n_1 - n_0$阶实正定对角阵D_2满足

$$
B_2 = U_2 \begin{bmatrix} D_2 & 0 \\ 0 & 0 \end{bmatrix} V_2^T. \tag{C.0.4}
$$

记

$$
B_1V_2 = \begin{bmatrix} B_{11} & B_{12} \end{bmatrix}, \tag{C.0.5}
$$

其中$B_{11} \in \mathbb{R}^{n_0\times(r+n_1-n_0)}$, $B_{12} \in \mathbb{R}^{n_0\times(n_0-n_1)}$. 由奇异值分解, 存在正交阵$U_3$, V_3和正定对角阵$D_3 \in \mathbb{R}^{(n_0-n_1)\times(n_0-n_1)}$满足

$$
D_1^{-1}B_{12} = U_3 \begin{bmatrix} 0 \\ D_3 \end{bmatrix} V_3^T.
$$

(注意: $\mathrm{rank}(D_1^{-1}B_{12}) = n_0 - n_1$由式(C.0.2), 式(C.0.4)和式(C.0.5)推出). 令

$$V = \mathrm{diag}(U_3^T, I_{n-n_0})V_1^T,\ M = \mathrm{diag}(D_3^{-1}, 0),$$

$$T = \begin{bmatrix} 0 & V_3^T \\ I_{r+n_1-n_0} & 0 \end{bmatrix} V_2^T,$$

$$N = U_1\mathrm{diag}(I_{n_0}, U_2)\left[\begin{array}{c|cc} D_1U_3\mathrm{diag}(I_{n_1}, D_3) & B_{11} & 0 \\ \hline 0 & D_2 & 0 \\ 0 & 0 & I \end{array}\right],$$

则

$$\begin{aligned} & N\begin{bmatrix} I_{n_1} & 0 \\ 0 & M \\ 0 & 0 \end{bmatrix}V \\ = \ & U_1\mathrm{diag}(I_{n_0}, U_2)\mathrm{diag}(D_1U_3, 0)V \\ = \ & U_1\mathrm{diag}(I_{n_0}, U_2)\mathrm{diag}(D_1U_3, 0)\mathrm{diag}(U_3^T, I_{n-n_0})V_1^T \\ = \ & U_1\mathrm{diag}(D_1, 0)V_1^T \\ = \ & E \end{aligned}$$

且

$$\begin{aligned} N\begin{bmatrix} 0_{n_1\times r} \\ I_r \\ 0 \end{bmatrix}T = \ & N\begin{bmatrix} 0_{n_1\times r} \\ I_r \\ 0 \end{bmatrix}\begin{bmatrix} 0 & V_3^T \\ I_{r+n_1-n_0} & 0 \end{bmatrix}V_2^T \\ = \ & U_1\mathrm{diag}(I_{n_0}, U_2)\left[\begin{array}{cc} B_{11} & D_1U_3\begin{bmatrix} 0 \\ D_3 \end{bmatrix}V_3^T \\ \hline D_2 & 0 \\ 0 & 0 \end{array}\right]V_2^T \\ = \ & U_1\mathrm{diag}(I_{n_0}, U_2)\left[\begin{array}{cc} B_{11} & B_{12} \\ \hline D_2 & 0 \\ 0 & 0 \end{array}\right]V_2^T \\ = \ & U_1\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \\ = \ & B, \end{aligned}$$

即式(3.4.3)成立.

附录 D 命题5.5.1至5.5.3的证明

命题5.5.1的证明 注意到$\|I_n+hA\|_2=\sqrt{\lambda_{\max}(I_n+hA)^*(I_n+hA)}$且矩阵的特征值是其元素的连续函数, 应用洛必达法则得

$$\begin{aligned}
\mu_{\|\cdot\|_2}(A)= & \lim_{h\to0^+}\frac{\|I_n+hA\|_2-1}{h}\\
= & \lim_{h\to0^+}\frac{\mathrm{d}\sqrt{\lambda_{\max}(I_n+hA)^*(I_n+hA)}}{\mathrm{d}h}\\
= & \frac{1}{2}\lim_{h\to0^+}\frac{\mathrm{d}\lambda_{\max}\left(I_n+h(A^*+A+hA^*A)\right)}{\mathrm{d}h}.
\end{aligned}$$

令$Y(h)=A^*+A+hA^*A,\lambda_1(h)\geqslant\lambda_2(h)\geqslant\cdots\geqslant\lambda_n(h)$为$Y(h)$的特征值, 则$I_n+hY(h)$的特征值为$1+h\lambda_1(h)\geqslant1+h\lambda_2(h)\geqslant\cdots\geqslant1+h\lambda_n(h)$, 因此

$$\begin{aligned}
\mu_{\|\cdot\|_2}(A)= & \frac{1}{2}\lim_{h\to0^+}\frac{\mathrm{d}\lambda_{\max}\left(I_n+hY(h)\right)}{\mathrm{d}h}\\
= & \frac{1}{2}\lim_{h\to0^+}\frac{\mathrm{d}\left(1+h\lambda_1(h)\right)}{\mathrm{d}h}\\
= & \frac{1}{2}\lim_{h\to0^+}\lambda_1(h)+\frac{1}{2}\lim_{h\to0^+}h\cdot\frac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h}\\
= & \frac{1}{2}\lambda_{\max}(A+A^*)+\frac{1}{2}\lim_{h\to0^+}h\cdot\frac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h}.
\end{aligned}$$

由$\mu_{\|\cdot\|_2}(A)$存在知$\lim\limits_{h\to0^+}h\dfrac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h}=k$存在, 只需证$k=0$. 若不然, 设$Y(h)$的特征多项式为

$$f(\lambda)=\det\left(\lambda I-Y(h)\right)=\lambda^n+a_{n-1}(h)\lambda^{n-1}+\cdots+a_1(h)\lambda+a_0(h),$$

其中$a_i(h)$为h的多项式, $i=0,1,\cdots,n-1$, 故

$$\lambda_1(h)^n+a_{n-1}(h)\lambda_1(h)^{n-1}+\cdots+a_1(h)\lambda_1(h)+a_0(h)=0.$$

上式两边对h求导得

$$\begin{aligned}
& n\lambda_1^{n-1}(h)\frac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h}+a'_{n-1}(h)\lambda_1^{n-1}(h)+(n-1)a_{n-1}(h)\lambda_1^{n-2}(h)\frac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h}\\
+ & \cdots+a'_1(h)\lambda_1(h)+a_1(h)\frac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h}+a'_0(h)=0.
\end{aligned}$$

令

$$c_0(h) = a'_{n-1}(h)\lambda_1^{n-1}(h) + \cdots + a'_1(h)\lambda_1(h) + a'_0(h),$$
$$d_0(h) = n\lambda_1^{n-1}(h) + (n-1)a_{n-1}(h)\lambda_1^{n-2}(h) + \cdots + a_1(h),$$

则$d_0(h)\dfrac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h} + c_0(h) = 0$, 于是

$$\frac{d_0(h)}{h}\left(h\frac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h}\right) + c_0(h) = 0.$$

由$\lim\limits_{h\to 0^+} c_0(h)$存在及

$$\lim_{h\to 0^+} h\frac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h} = k \neq 0$$

知$\lim\limits_{h\to 0^+}\dfrac{d_0(h)}{h}$存在, 故$\lim\limits_{h\to 0^+} d'_0(h)$存在. 注意到

$$d'_0(h) = \frac{d_1(h)}{h}\left(h\frac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h}\right) + c_1(h),$$

其中

$$d_1(h) = n(n-1)\lambda_1^{n-2}(h) + (n-1)(n-2)a_{n-1}(h)\lambda_1^{n-3}(h) + \cdots + 2a_2(h),$$
$$c_1(h) = (n-1)a'_{n-1}(h)\lambda_1^{n-2}(h) + \cdots + 2a'_2(h)\lambda_1(h) + a'_1(h).$$

类似推出$\lim\limits_{h\to 0^+} d'_0(h)$存在的过程易得$\lim\limits_{h\to 0^+} d'_1(h)$存在. 进而$\lim\limits_{h\to 0^+} d'_{n-2}(h)$存在. 注意到

$$d'_{n-2}(h) = n!\frac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h} + g(h) \quad (g(h)\text{是关于}h\text{的多项式}),$$

我们推出$\lim\limits_{h\to 0^+}\dfrac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h}$存在, 故$k = \lim\limits_{h\to 0^+} h\dfrac{\mathrm{d}\lambda_1(h)}{\mathrm{d}h} = 0$, 矛盾, 因而$k = 0$. 从而$\mu_{\|\cdot\|_2}(A) = \frac{1}{2}\lambda_{\max}(A + A^*)$.

命题5.5.2的证明 由$\|A\|_1 = \max\limits_j \sum\limits_{i=1}^n |a_{ij}|$, $\forall A = [a_{ij}] \in \mathbb{F}^{n\times n}$及洛必达法则有

$$\begin{aligned}
\mu_{\|\cdot\|_1}(A) = &\lim_{h\to 0^+}\frac{\|I_n + hA\|_1 - 1}{h} \\
= &\lim_{h\to 0^+}\frac{\max\limits_{1\leqslant j\leqslant n}\left\{|1 + ha_{jj}| + \sum\limits_{\substack{i=1\\ i\neq j}}^n |ha_{ij}|\right\} - 1}{h} \\
= &\max_{1\leqslant j\leqslant n}\lim_{h\to 0^+}\frac{|1 + ha_{jj}| + h\sum\limits_{\substack{i=1\\ i\neq j}}^n |a_{ij}| - 1}{h}
\end{aligned}$$

$$= \max_{1\leqslant j\leqslant n}\lim_{h\to 0^+}\left(\frac{\sqrt{(1+h\mathrm{Re}(a_{jj}))^2+h^2\left(\mathrm{I_m}(a_{jj})\right)^2}-1}{h}+\sum_{\substack{i=1\\i\neq j}}^{n}|a_{ij}|\right)$$

$$= \max_{1\leqslant j\leqslant n}\lim_{h\to 0^+}\left(\frac{2(1+h\mathrm{Re}(a_{jj}))\mathrm{Re}(a_{jj})+2h\left(\mathrm{I_m}(a_{jj})\right)^2}{2\sqrt{(1+h\mathrm{Re}(a_{jj}))^2+h^2\left(\mathrm{I_m}(a_{jj})\right)^2}}+\sum_{\substack{i=1\\i\neq j}}^{n}|a_{ij}|\right)$$

$$= \max_{1\leqslant j\leqslant n}\left(\mathrm{Re}(a_{jj})+\sum_{\substack{i=1\\i\neq j}}^{n}|a_{ij}|\right).$$

命题5.5.3的证明 类似于命题5.5.2的证明.

附录 E 定理8.3.1的证明

引理 E.0.1 设线性空间V的子空间$P_1,\cdots,P_k$满足$P_1\supset\cdots\supset P_k$且

$$\dim P_i\geqslant k-i+1,\ i=1,2,\cdots,k,$$

$\xi_i\in P_i$, $i=1,2,\cdots,k-1$线性无关, $M=\operatorname{span}\{\xi_1,\cdots,\xi_{k-1}\}$, 则存在$x\in P_1\setminus M$使得$M\oplus\operatorname{span}\{x\}$有基$\eta_1,\cdots,\eta_k$且$\eta_i\in P_i,i=1,2,\cdots,k$.

证明 对k应用数学归纳法. $k=1$时显然(此时$M=\{0\}$). 假设小于k时结论成立.

当k时, 令$N=\operatorname{span}\{\xi_2,\cdots,\xi_{k-1}\}$(当$k=2$时令$N=\{0\}$), 由归纳假设, 存在$y\in P_2\setminus N$使得$N\oplus\operatorname{span}\{y\}$有基$\xi_2',\cdots,\xi_k'$且$\xi_j'\in P_j,j=2,\cdots,k$.

若$y\notin M$, 则$\xi_1\notin N\oplus\operatorname{span}\{y\}$(否则, $M\subset N\oplus\operatorname{span}\{y\}$, 这和$\dim M=\dim(N\oplus\operatorname{span}\{y\})$一起推出$M=N\oplus\operatorname{span}\{y\}$, 与$y\notin M$矛盾), 因而$M\oplus\operatorname{span}\{y\}$有基$\xi_1,\xi_2',\cdots,\xi_k'$且$\xi_1\in P_1,\xi_j'\in P_j,j=2,\cdots,k$; 若$y\in M$, 则$y=\sum\limits_{i=1}^{k-1}\alpha_i\xi_i$, 由$y\notin N$推出$\alpha_1\neq 0$, 故

$$N\oplus\operatorname{span}\{y\}=M,$$

于是M有基$\xi_2',\cdots,\xi_k'$且$\xi_j'\in P_j,j=2,\cdots,k$. 再由$\dim P_1\geqslant k>k-1=\dim M$知存在$\xi_1'\in P_1\setminus M$使得$M\oplus\operatorname{span}\{\xi_1'\}$有基$\xi_1',\xi_2',\cdots,\xi_k'$且$\xi_i'\in P_i,i=1,2,\cdots,k$. □

引理 E.0.2 设$1\leqslant i_1<\cdots<i_k\leqslant n,M_1\subset M_2\subset\cdots\subset M_k$和$N_1\supset N_2\supset\cdots\supset N_k$均是$n$维线性空间$V$的子空间, 且$\dim M_p=i_p,\dim N_p=n-i_p+1,p=1,2,\cdots,k$,则存在$V$的子空间$M$使得$M$有基$\xi_p\in M_p,p=1,2,\cdots,k$和基$\eta_p\in N_p,p=1,2,\cdots,k$.

证明 对k应用数学归纳法. 当$k=1$时, 由$\dim M_1+\dim N_1=n+1$知$M_1\cap N_1\neq\{0\}$, 于是存在$0\neq\xi_1=\eta_1\in M_1\cap N_1$, 即结论成立. 假设当$k-$

1时定理成立, 即存在V的子空间M_0使得M_0有基$\xi_j \in M_j, j = 1,2,\cdots,k-1$和基$\eta_j \in N_j, j = 1,2,\cdots,k-1$.

当k时, 令$P_j = N_j \cap M_k, j = 1,2,\cdots,k$, 则$P_1 \supset P_2 \supset \cdots \supset P_k$. 注意到$i_k - i_j \geqslant k-j$, 由

$$\dim M_k + \dim N_j = i_k + n - i_j + 1 \geqslant n + k - j + 1$$

得

$$\begin{aligned}\dim P_j = &\ \dim(N_j \cap M_k) = \dim M_k + \dim N_j - \dim(M_k + N_j)\\ \geqslant &\ n + k - j + 1 - n = k - j + 1, j = 1,2,\cdots,k.\end{aligned}$$

另一方面, 由$\eta_j \in M_0 \subset M_k$及$\eta_j \in N_j, j = 1,2,\cdots,k-1$推出$\eta_j \in P_j, j = 1,2,\cdots,k-1$. 使用引理E.0.1知存在$x \in P_1 \backslash M_0$使得$M_0 \oplus \operatorname{span}\{x\}$有基$\eta'_s \in P_s \subset N_s, s = 1,2,\cdots,k$. 另外, $M_0 \oplus \operatorname{span}\{x\}$有基$\xi_1,\cdots,\xi_{k-1},x$且$x \in M_k, \xi_j \in M_j, j = 1,2,\cdots,k-1$. □

推论 E.0.1 若在引理E.0.2的前提条件下附加假设V是酉空间, 则存在V的子空间M使得M有标准正交基$\xi_i \in M_i, i = 1,2,\cdots,k$及标准正交基$\eta_i \in N_i, i = 1,2,\cdots,k$.

证明 由引理E.0.2知, 存在V的子空间M使得M有基

$$x_i \in M_i,\ i = 1,2,\cdots,k$$

及基

$$y_i \in N_i,\ i = 1,2,\cdots,k.$$

再分别将$x_1,\cdots,x_k$及$y_1,\cdots,y_k$标准正交化后可得M的标准正交基$\xi_i \in M_i$, $i = 1,\cdots,k$及标准正交基$\eta_i \in N_i$, $i = 1,2,\cdots,k$. □

引理 E.0.3 设$H \in \mathbb{C}^{n\times n}$满足$H^* = H$, φ及$\widetilde{\varphi}$同前面定义, 若$\mathbb{C}^n$中的两个标准正交的向量组$x_1,\cdots,x_k$及$y_1,\cdots,y_k$满足

$$\operatorname{span}\{x_1,\cdots,x_k\} = \operatorname{span}\{y_1,\cdots,y_k\},$$

则

$$\widetilde{\varphi}(x_1,\cdots,x_k) = \widetilde{\varphi}(y_1,\cdots,y_k).$$

证明 易见, 存在酉矩阵U满足

$$[y_1,\cdots,y_k] = [x_1,\cdots,x_k]U,$$

于是矩阵$[x_1,\cdots,x_k]^*H[x_1,\cdots,x_k]$与矩阵$[y_1,\cdots,y_k]^*H[y_1,\cdots,y_k]$有相同的特征值, 故$\widetilde{\varphi}(x_1,\cdots,x_k) = \widetilde{\varphi}(y_1,\cdots,y_k)$. □

引理 E.0.4　设V是复n维酉空间, V的子空间N_1, $\cdots$, N_n满足

$$N_1 \subset \cdots \subset N_n,\ \dim N_i = i,\ i = 1, 2, \cdots, n.$$

正整数i_1, $\cdots$, i_k满足$i_1 \leqslant \cdots \leqslant i_k \leqslant n$且$i_p \geqslant p$, $p = 1$, 2, $\cdots$, k. 若$z_p \in N_{i_p}$, $p = 1$, 2, $\cdots$, k满足z_1, $\cdots$, z_k标准正交, 则存在$y_p \in N_{i'_p}$, $p = 1$, 2, $\cdots$, k满足y_1, $\cdots$, y_k标准正交且

$$\text{span}\{z_1, \cdots, z_k\} = \text{span}\{y_1, \cdots, y_k\}.$$

证明　对k应用数学归纳法. 当$k = 1$时结论显然成立. 假设$k - 1(k > 1)$时结论成立. 下面往证当k时结论成立.

(i) $i_k > i_{k-1}$. 由归纳假设, 存在$y_p \in N_{i'_p}, p = 1, 2, \cdots, k-1$使得$y_1$, $\cdots$, y_{k-1} 标准正交, 且$\text{span}\{y_1, \cdots, y_{k-1}\} = \text{span}\{z_1, \cdots, z_{k-1}\}$, 于是

$$\text{span}\{y_1, \cdots, y_{k-1}, z_k\} = \text{span}\{z_1, \cdots, z_k\}.$$

再由Schmidt正交化知, 存在向量$y_k \in N_{i_k}$使得

$$\text{span}\{y_1, \cdots, y_k\} = \text{span}\{z_1, \cdots, z_k\}.$$

由推论8.3.1知$i_k = i'_k$, 从而结论成立.

(ii) $i_k = i_{k-1}$. 此时一定存在正整数m满足

$$i_1 \leqslant \cdots \leqslant i_{m-1} < i_m = \cdots = i_k.$$

令$W_m = \text{span}\{z_m, \cdots, z_k\}$, 则$N_{i_k-1} \cup W_m \subset N_{i_k}$, 于是

$$\dim(N_{i_k-1} \cap W_m) \geqslant k - m.$$

设w_m, $\cdots$, w_{k-1}是$N_{i_k-1} \cap W_m$的标准正交基, 并在W_m中选择w_k使得w_k, w_m, $\cdots$, w_{k-1}标准正交. 使用$z_1, \cdots, z_k$标准正交推出z_1, $\cdots$, z_{m-1}, w_m, $\cdots$, w_k标准正交且

$$\text{span}\{z_1, \cdots, z_k\} = \text{span}\{z_1, \cdots, z_{m-1}, w_m, \cdots, w_k\}.$$

对正整数序列i_1, $\cdots$, i_{m-1}, $i_k - 1$, $\cdots$, $i_k - 1$, i_k(其中$i_k - 1$为$k - m$个)和向量序列z_1, $\cdots$, z_{m-1}, w_m, $\cdots$, w_k应用(i)容易推出要证结论. □

引理 E.0.5　设$H \in \mathbb{C}^{n\times n}$满足$H^* = H$. $\lambda_1 \geqslant \cdots \geqslant \lambda_n$是$H$的特征值. 正整数$i_1, \cdots, i_k$满足$i_1 < \cdots < i_k \leqslant n$.

(i) 若$\mathbb{C}^n$的子空间$V_1, \cdots, V_k$满足$V_1 \subset \cdots \subset V_k$且$\dim V_p = i_p, p = 1, \cdots, k$, 则

$$\inf_{\substack{x_p \in V_p \\ \{x_p\} o.n.}} \widetilde{\varphi}(x_1, \cdots, x_k) \leqslant \varphi(\lambda_{i_1}, \cdots, \lambda_{i_k}),$$

其中$\widetilde{\varphi}(x_1, \cdots, x_k)$同前面定义.

(ii) $\sup\limits_{\substack{M_1 \subset \cdots \subset M_k \\ \dim M_p = i_p}} \inf\limits_{\substack{x_p \in M_p \\ \{x_p\} o.n.}} \widetilde{\varphi}(x_1, \cdots, x_k) \leqslant \varphi(\lambda_{i_1}, \cdots, \lambda_{i_k}).$

证明 设$Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}$为酉阵且$H = Q\mathrm{diag}(\lambda_1, \cdots, \lambda_n)Q^*$, 记

$$N_p = \mathrm{span}\left\{q_{i_p}, \cdots, q_n\right\},$$

则$N_1 \supset \cdots \supset N_k$且$\dim N_p = n - i_p + 1, p = 1, 2, \cdots, k$. 注意到$V_1 \subset \cdots \subset V_k$满足$\dim V_p = i_p$, $p = 1, 2, \cdots, k$. 由推论E.0.1知, 存在$\mathbb{C}^n$的子空间M使得M有标准正交基$\xi_p \in V_p$, $p = 1, 2, \cdots, k$及标准正交基$\eta_p \in N_p, p = 1, 2, \cdots, k$, 从而存在酉阵$U$使得$[\xi_1 \cdots \xi_k]U = [\eta_1 \cdots \eta_k]$. 使用引理E.0.3推出

$$\inf_{\substack{x_p \in V_p \\ \{x_p\} o.n.}} \widetilde{\varphi}(x_1, \cdots, x_k) \leqslant \widetilde{\varphi}(\xi_1, \cdots, \xi_k) = \widetilde{\varphi}(\eta_1, \cdots, \eta_k). \tag{E.0.1}$$

设$\alpha_1, \cdots, \alpha_k$是矩阵$[\eta_1 \cdots \eta_k]^* H [\eta_1 \cdots \eta_k]$的特征值, 则由定理8.1.4(i)得

$$\begin{aligned}
\alpha_p &= \min_{W \in \mathbb{C}^{k \times (k-p+1)}_{k-p+1}} \max_{\substack{z \in \mathcal{R}(W) \\ \|z\|_2 = 1}} z^*[\eta_1 \cdots \eta_k]^* H [\eta_1 \cdots \eta_k] z \\
&= \min_{\substack{S \in \mathbb{C}^{n \times (k-p+1)}_{k-p+1} \\ \mathcal{R}(S) \subseteq \mathrm{span}\{\eta_1, \cdots, \eta_k\}}} \max_{\substack{z \in \mathcal{R}(S) \\ \|z\|_2 = 1}} z^* H z \\
&\leqslant \max_{\substack{z \in \mathrm{span}\{\eta_p, \cdots, \eta_k\} \\ \|z\|_2 = 1}} z^* H z \\
&\leqslant \max_{\substack{z \in N_p \\ \|z\|_2 = 1}} z^* H z \\
&= \lambda_{i_p}.
\end{aligned}$$

使用$\varphi(t_1, \cdots, t_k)$关于每个变量递增推出

$$\widetilde{\varphi}(\eta_1, \cdots, \eta_k) = \varphi(\alpha_1, \cdots, \alpha_k) \leqslant \varphi(\lambda_{i_1}, \cdots, \lambda_{i_k}).$$

这和式(E.0.1)一起推出(i)成立, 从而(ii)成立. □

推论 E.0.2 设$H \in \mathbb{C}^{n \times n}$满足$H^* = H$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$是$H$的特征值, 正整数$i_1, \cdots, i_k$满足$i_1 \leqslant \cdots \leqslant i_k \leqslant n$, 且$i_p \geqslant p, p = 1, 2, \cdots, k$, φ及$\widetilde{\varphi}$同前面定义, 则

$$\sup_{\substack{M_1 \subseteq \cdots \subseteq M_k \\ \dim M_p = i_p}} \inf_{\substack{x_p \in M_p \\ \{x_p\} o.n.}} \widetilde{\varphi}(x_1, \cdots, x_k) \leqslant \varphi(\lambda_{i_1'}, \cdots, \lambda_{i_k'}).$$

证明 若$\mathbb{C}^n$的子空间$M_1,\cdots,M_k$满足$M_1\subseteq\cdots\subseteq M_k$及$\dim M_p=i_p$, $p=1,2,\cdots,k$, 则存在$\mathbb{C}^n$的子空间$M_1',\cdots,M_k'$满足$M_1'\subset\cdots\subset M_k'$且$M_{p'}\subseteq M_p$, $\dim M_p'=i_p'$, $p=1,2,\cdots,k$, 于是

$$\inf_{\substack{x_p\in M_p\\ \{x_p\}o.n.}}\widetilde{\varphi}(x_1,\cdots,x_k)\leqslant\inf_{\substack{x_p\in M_p'\\ \{x_p\}o.n.}}\widetilde{\varphi}(x_1,\cdots,x_k).$$

由引理E.0.5得

$$\inf_{\substack{x_p\in M_p\\ \{x_p\}o.n.}}\widetilde{\varphi}(x_1,\cdots,x_k)\leqslant\varphi(\lambda_{i_1'},\cdots,\lambda_{i_k'}),$$

从而结论得证. □

引理 E.0.6 设$H\in\mathbb{C}^{n\times n}$满足$H^*=H$, $\lambda_1\geqslant\cdots\geqslant\lambda_n$是$H$的特征值, 对$i=1,2,\cdots,n$, q_i是H的属于λ_i的特征向量, 并且$q_1,\cdots,q_n$标准正交, 正整数$i_1,\cdots,i_k$满足$i_1\leqslant\cdots\leqslant i_k\leqslant n$且$i_p\geqslant p$, $p=1,2,\cdots,k$, φ及$\widetilde{\varphi}$同前面定义, 记$Q_j=\mathrm{span}\{q_1,\cdots,q_j\}$, $j=1,2,\cdots,n$, 则

(i) $\displaystyle\inf_{\substack{x_p\in Q_{i_p'}\\ \{x_p\}o.n.}}\widetilde{\varphi}(x_1,\cdots,x_k)\geqslant\varphi(\lambda_{i_1'},\cdots,\lambda_{i_k'})$;

(ii) $\displaystyle\inf_{\substack{x_p\in Q_{i_p}\\ \{x_p\}o.n.}}\widetilde{\varphi}(x_1,\cdots,x_k)\geqslant\varphi(\lambda_{i_1'},\cdots,\lambda_{i_k'})$.

证明 (i) 对任意满足$x_1,\cdots,x_k$标准正交的$x_p\in Q_{i_p'}$, $p=1,2,\cdots,k$. 由定理8.1.4(ii)知

$$\begin{aligned}\mu_p&=\max_{M\in\mathbb{C}_p^{k\times p}}\min_{\substack{y\in\mathcal{R}(M)\\ \|y\|_2=1}}y^*[x_1\cdots x_k]^*H[x_1\cdots x_k]y\\
&=\max_{\substack{W\in\mathbb{C}_p^{n\times p}\\ \mathcal{R}(W)\subseteq\mathrm{span}\{x_1,\cdots,x_k\}}}\min_{\substack{z\in\mathcal{R}(W)\\ \|z\|_2=1}}z^*Hz\\
&\geqslant\min_{\substack{z\in\mathrm{span}\{x_1,\cdots,x_k\}\\ \|y\|_2=1}}z^*Hz\\
&\geqslant\min_{\substack{x\in Q_{i_p'}\\ \|x\|_2=1}}x^*Hx\\
&=\lambda_{i_p'},p=1,2,\cdots,k.\end{aligned}$$

使用φ关于每个变量递增推出

$$\widetilde{\varphi}(x_1,\cdots,x_k)=\varphi(\mu_1,\cdots,\mu_k)\geqslant\varphi(\lambda_{i_1'},\cdots,\lambda_{i_k'}).$$

从而结论得证.

(ii) 令

$$S = \left\{(x_1, \cdots, x_k) | x_p \in Q_{i_p}, p = 1, 2, \cdots, k, \{x_p\}\, o.n.\right\},$$

$$S' = \left\{(x_1, \cdots, x_k) | x_p \in Q_{i'_p}, p = 1, 2, \cdots, k, \{x_p\}\, o.n.\right\},$$

$$\begin{aligned} S'_0 = \ & \Big\{(x_1, \cdots, x_k) : (x_1, \cdots, x_k) \in S' \text{且存在} (z_1, \cdots, z_k) \in S \\ & \text{使得span}\,\{x_1, \cdots, x_k\} = \text{span}\,\{z_1, \cdots, z_k\}\Big\}, \end{aligned}$$

则$S'_0 \subseteq S'$, 从而由引理E.0.4及引理E.0.3得

$$\begin{aligned} \inf_{\substack{x_p \in Q_{i_p} \\ \{x_p\} o.n.}} \widetilde{\varphi}(x_1, \cdots, x_k) = & \inf_{(x_1, \cdots, x_k) \in S'_0} \widetilde{\varphi}(x_1, \cdots, x_k) \\ \geqslant & \inf_{(x_1, \cdots, x_k) \in S'} \widetilde{\varphi}(x_1, \cdots, x_k) \\ = & \inf_{\substack{x_p \in Q_{i'_p} \\ \{x_p\} o.n.}} \widetilde{\varphi}(x_1, \cdots, x_k). \end{aligned}$$

使用(i)推出(ii)成立. □

推论 E.0.3 设$H \in \mathbb{C}^{n \times n}$满足$H^* = H$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$是$H$的特征值, 正整数$i_1, \cdots, i_k$满足$i_1 \leqslant \cdots \leqslant i_k \leqslant n$, 且$i_p \geqslant p$, $p = 1, 2, \cdots, k$, φ及$\widetilde{\varphi}$同前面定义, 则

$$\sup_{\substack{M_1 \subseteq \cdots \subseteq M_k \\ \dim M_p = i_p}} \inf_{\substack{x_p \in M_p \\ \{x_p\} o.n.}} \widetilde{\varphi}(x_1, \cdots, x_k) \geqslant \varphi(\lambda_{i'_1}, \cdots, \lambda_{i'_k}).$$

组合推论E.0.2和推论E.0.3得到定理8.3.1.

附录 F 定理11.5.1的证明

令

$$\underline{\mu}=\min_{\lambda\in[0,T]}\lambda_{\max}\Phi(\lambda),\overline{\mu}=\max_{\lambda\in[0,T]}\lambda_{\max}\Phi(\lambda),$$

则$\overline{\mu}\geqslant\underline{\mu}>0$. 对于待定的正整数$N$, 令

$$T_i=\frac{i}{N}T,\mu_i=\lambda_{\max}\Phi(T_i),i=0,1,\cdots,N,$$

则$\underline{\mu}\leqslant\mu_i\leqslant\overline{\mu},i=0,1,\cdots,N.$

设η_i为$\Phi(T_i)$的属于特征值μ_i的单位特征向量, 并满足

$$\eta_i^T\eta_{i-1}\geqslant 0,i=1,2,\cdots,N,$$

则

$$\Phi(T_i)\eta_i=\mu_i\eta_i,i=0,1,\cdots,N.$$

定义连续函数$\eta:[0,T]\to\mathbb{R}^n$如下:

$$\eta(\lambda)=\frac{T_i-\lambda}{T_i-T_{i-1}}\eta_{i-1}+\frac{\lambda-T_{i-1}}{T_i-T_{i-1}}\eta_i,\forall\lambda\in[T_{i-1},T_i],i=1,2,\cdots,N.$$

显然(ii)成立, 下面往证(i)成立.

任取$\lambda\in[0,T]$, 则存在j使得$\lambda\in[T_{j-1},T_j]$, 于是

$$\lambda=\alpha T_{j-1}+(1-\alpha)T_j,$$

其中

$$\alpha=\frac{T_j-\lambda}{T_j-T_{j-1}},$$

从而

$$\Phi(\lambda)=\Phi(\alpha T_{j-1}+(1-\alpha)T_j)=\alpha\Phi(T_{j-1})+(1-\alpha)\Phi(T_j)-\alpha(1-\alpha)\frac{T^2}{N^2}X,$$

故

$$\begin{aligned}
&\eta^T(\lambda)\Phi(\lambda)\eta(\lambda)\\
=\ &[\alpha\eta_{j-1}^T+(1-\alpha)\eta_j^T][\alpha\Phi(T_{j-1})+(1-\alpha)\Phi(T_j)][\alpha\eta_{j-1}+(1-\alpha)\eta_j]\\
&+\triangle_1(N)\\
=\ &\alpha^3\mu_{j-1}+2\alpha^2(1-\alpha)\mu_{j-1}\eta_{j-1}^T\eta_j+2\alpha(1-\alpha)^2\mu_j\eta_{j-1}^T\eta_j\\
&+\alpha^2(1-\alpha)\eta_{j-1}^T\Phi(T_j)\eta_{j-1}+\alpha(1-\alpha)^2\eta_j^T\Phi(T_{j-1})\eta_j\\
&+\triangle_1(N)+(1-\alpha)^3\mu_j\\
\geqslant\ &\alpha^3\mu_{j-1}+(1-\alpha)^3\mu_j+\alpha^2(1-\alpha)\eta_{j-1}^T\Phi(T_j)\eta_{j-1}\\
&+\alpha(1-\alpha)^2\eta_j^T\Phi(T_{j-1})\eta_j+\triangle_1(N).
\end{aligned}$$

使用$\Phi(T_j)=\Phi(T_{j-1}+\frac{T}{N})$及$\Phi(T_{j-1})=\Phi(T_j+\frac{T}{N})$推出

$$\begin{aligned}
\eta^T(\lambda)\Phi\eta(\lambda)\geqslant\ &\alpha^2\mu_{j-1}+(1-\alpha)^2\mu_j+\triangle(N)\\
=\ &(\mu_{j-1}+\mu_j)\alpha^2-2\mu_j\alpha+\mu_j+\triangle(N)\\
\geqslant\ &\frac{\mu_{j-1}\mu_j}{\mu_{j-1}+\mu_j}+\triangle(N)\\
\geqslant\ &\frac{\mu^2}{2\overline{\mu}}+\triangle(N).
\end{aligned}$$

由$\lim\limits_{N\to+\infty}\triangle(N)=\lim\limits_{N\to+\infty}\triangle_1(N)=0$知当$N$充分大时有$\frac{\mu^2}{2\overline{\mu}}+\triangle(N)>0$, 因而(i)成立.

[General Information]
书名=系统与控制中的矩阵理论
作者=张显，仲光苹，高翔宇编著
页数=206
SS号=13024534
DX号=
出版日期=2011.08
出版社=黑龙江大学出版社